

From: Rao, Manjunath N.
Sent: Thursday, November 07, 2002 3:37 PM
To: STIC-Biotech/ChemLib
Subject: Sequence search request for 09/917,378

From: Manjunath N. Rao
Art Unit 1652, Room 10A11
Mail Box in Room 10D 01
Phone: 306-5681

Date: 11-7-02

Please search the following as soon as possible for application with serial number **09/917,378**

SEQ ID NO: 2, against all commercial nucleic acid databases including issued patents database and pending application database and provide a **print of all results**.

SEQ ID NO:1, 3, 4, and 5, against all commercial protein databases including issued patents database and pending application database and provide a **print of all results**.

If you have any questions please call me at the above phone number.

Thanks

Manjunath N. Rao, Ph.D.
Biotechnology Patent Examiner
Art Unit 1652, Room 10A11
Crystal Mall 1, USPTO.

Point of Contact:
Toby Port
Technical Info. Specialist
CM1 6A04
703-308-3534

*None have a
100% match.
No w-f or seq. match
except for Seq ID 5 alone.*

Searcher: _____
Phone: _____
Location: _____
Date Picked Up: _____
Date Completed: _____
Searcher Prep/Review: _____
Clerical: _____
Online time: _____

TYPE OF SEARCH:
NA Sequences: _____
AA Sequences: _____
Structures: _____
Bibliographic: _____
Litigation: _____
Full text: _____
Patent Family: _____
Other: _____

VENDOR/COST (where applic.)
STN: _____
DIALOG: _____
Questel/Orbit: _____
DRLink: _____
Lexis/Nexis: _____
Sequence Sys.: _____
WWW/Internet: _____
Other (specify): _____

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 13, 2002, 10:49:26 : Search time 68.4267 Seconds
(without alignments)
1483.879 Million cell updates/sec

Title: us-09-917-378-1

Perfect score: 4122

Sequence: 1 MGLVRPARAFVATAAGTAV.....GFNGSYSGTNTAPTTLCTAS 762

Scoring table: BIOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A_Geneseq_101002:*

1: /SID52/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:*

2: /SID52/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:*

3: /SID52/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:*

4: /SID52/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:*

5: /SID52/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:*

6: /SID52/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:*

7: /SID52/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:*

8: /SID52/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:*

9: /SID52/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:*

10: /SID52/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:*

11: /SID52/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:*

12: /SID52/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:*

13: /SID52/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:*

14: /SID52/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:*

15: /SID52/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:*

16: /SID52/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:*

17: /SID52/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:*

18: /SID52/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:*

19: /SID52/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:*

20: /SID52/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:*

21: /SID52/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:*

22: /SID52/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:*

23: /SID52/gcgdata/geneseq/geneseq-emb1/AA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	841.5	20.4	666	19 AAW34992	Thermotoga neapoli
2	813	19.7	680	18 AAW34564	Thermotoga maritim
3	813	19.7	680	19 AAW35005	Thermotoga maritim
4	812	19.7	663	19 AAW49872	Thermotoga maritim
5	806	19.6	680	19 AAW49868	Thermotoga maritim
6	698.5	16.9	562	21 AAY69508	Acidothermus cellu
7	696.5	16.9	562	23 AAY79549	A. cellulolyticus
8	676.5	16.4	521	22 AAB48786	Acidothermus cellu
9	674.5	16.4	521	17 AAR89927	A. cellulolyticus
10	670.5	16.3	521	22 AAB48787	Acidothermus cellu

11	668.5	16.2	521	22 AAB48788	Acidothermus cellu
12	624	15.1	1751	20 AAV13493	Truncated cellulase
13	624	15.1	1751	23 AAE16324	Active cellulase p
14	589.5	14.3	1426	23 AAY13492	Truncated cellulase
15	589.5	14.3	1426	23 AAE16323	Active cellulase p
16	552	13.4	431	21 AAG27501	Arabidopsis thaliana
17	552	13.4	431	21 AAG42172	Arabidopsis thaliana
18	552	13.4	431	23 ABB93972	Herbicidally activ
19	552	13.4	442	21 AAG42171	Arabidopsis thaliana
20	552	13.4	443	21 AAG27500	Arabidopsis thaliana
21	544.5	13.2	416	22 AAG79297	Amino acid sequenc
22	541	13.1	408	23 ABB92275	Herbicidally activ
23	526.5	12.8	431	23 ABB90809	Herbicidally activ
24	519	12.6	414	23 ABB92274	Herbicidally activ
25	498	12.1	448	23 ABB93325	Herbicidally activ
26	495	12.0	431	23 ABB93333	Herbicidally activ
27	494	12.0	427	21 AAY93441	Amino acid sequenc
28	488	11.8	551	18 AAW18790	Corrected bacillus
29	484	11.7	1853	19 AAW43108	C. thermocellum ce
30	479	11.6	531	18 AAW15238	Scatfoldin protein
31	477	11.6	375	21 AAG27502	Arabidopsis thaliana
32	477	11.6	375	21 AAG42173	Arabidopsis thaliana
33	472	11.5	351	23 ABB92490	Herbicidally activ
34	469.5	11.4	616	20 AAY13494	Truncated cellulase
35	469.5	11.4	616	23 AAE16325	Active cellulase h
36	461.5	11.2	415	22 AAE00417	Lycopodium escul
37	451.5	11.0	499	14 AAR42122	NR-1 cellulase. R
38	448.5	10.9	782	12 AAR15625	Cellulase AE-1. A
39	446.5	10.8	476	21 AAY54123	A mammalian linker
40	439.5	10.7	677	23 AAY98063	Bacillus subtilis
41	438	10.6	700	12 AAR13227	Novel endoglucanase
42	426.5	10.3	369	20 AAV14150	Protein encoded by
43	416.5	10.1	403	23 ABB91820	Herbicidally activ
44	415.5	10.1	493	20 AAY28850	Pectate lyase-link
45	415.5	10.1	493	20 AAY43218	Pectate lyase CMO

ALIGNMENTS

RESULT 1

AAW34992

ID AAW34992 standard; Protein; 666 AA.

XX

AC AAW34992:

XX

DT 21-MAY-1998 (first entry)

XX

DE Thermotoga neapolitana endoglucanase.

DE

XX

DE Endoglucanase; cellulase; carboxymethylcellulose; cellulose;

KW biomass; beta-1,4-glycosidic bond; hydrolysis; saccharification;

KW thermostable enzyme; thermophilic; glycosidase.

XX

OS Thermotoga neapolitana (Clone 56Gp1).

XX

PN WO9744361-A1.

XX

PD 27-NOV-1997.

XX

PF 22-MAY-1997; 97WO-US08793.

XX

PR 22-MAY-1996; 96US-0651572.

XX

PA (RECO-) RECOMBINANT BIOTCATALYSIS INC.

XX

P1 Lam DE, Mathur EJ:

XX

DR WPI: 1998-018435/02.

DR N-PSDB; AAT94200.

XX

PT Endoglucanase(s), preferably from archaeal bacterium, APPI1 la -

PT useful to degrade carboxymethylcellulose and hydrolyse of

PT beta-1,4-glycosidic bonds in cellulose
 PS Claim 1: Fig 1B; 164pp; English.
 XX

This protein comprises an endoglucanase of *Thermotoga neapolitana* (Clone 56GPI) that is capable of degrading carboxymethylcellulose and of hydrolysing the beta-1,4-glycosidic bonds in cellulose. It has homology to an endoglucanase of *archaeobacterium* AEP11a (see AAW34985). It can be produced from native cells or from recombinant host cells, especially prokaryotic host cells transformed with a plasmid or virus-derived vector including the endoglucanase DNA (see AAT94200). 24 Endoglucanases (see AAW34986-W35008) are claimed. They can be used to degrade cellulose for the conversion of plant biomass into fuels and chemicals, for use in detergents, textiles, animal feed, waste treatment, and in the fruit juice and brewing industries for the clarification and extraction of juices.

XX
 SO Sequence 666 AA:

Query Match 20.4%; Score 841.5; DB 19; Length 666;
 Best Local Similarity 43.1%; Pred. No. 1.8e-40;
 Matches 175; Conservative 65; Mismatches 123; Indels 43; Gaps 13;

QY 29 ITMPSATAAPAGVVTASGQFVNLGLPYRYGGTNNYYLSYQSHADVDDVLAKAQAANLSV 88
 DB 11 VTLPVLFANSDFVKVENGFRFLNCEFRFGVSNYYMYHYKSNRMIDSVLESAAKAMGVK 70
 QY 89 IRTWGFIDIGSLDGSVPTIDGNKNGFYFOYWDPSGCA---PAYNDGPTGLOGLDYATASA 145
 DB 71 LRHWG-----LDGSYCRDKN-----TYMHPAGVFGLPCTNAQDGFERLDYTVAKA 119
 QY 146 AARGLRVIVLTNDMKFEGMDQYDKWGLPYHDNFTDPTQOAYKNMNHLLNRVNSI 205
 DB 120 KELGKLLVVLNNWDFGGMQYVRFVFCGTHDDFYRNEKIKEYKYVSEFLNRVNTY 179
 QY 206 TGVTYKNDPTIFAWELANPRCVSGTCLPTSGTCTQATIVNWDQMSATVKSIDPNHMY 265
 DB 180 TGVPREPTIFAWELANPRC-----ETDKSGN-----TLVWEVEMSAYIKSLDPNHLVA 231
 QY 266 VGDEGF---YIGSTCGS---WPYNDPSGVDNNALLRVKNDPFTGYHLYPNYNGON--- 316
 DB 232 VGDEGFFNNYEGFRPYGGAWAYNCWS-GVDWKLLLEIETVDFGTFHLYPSHWGVSPE 290
 QY 317 -ADMGTQWIKDHIAAAAIKPTILKEFG--WQTP-DRDSVYQVTTQVTRTNGEAGWNFW 372
 DB 291 YAQWGAKWIEDIIKIAKEGKPVVLEEYGIKPSAPVNRVATYKLMNDLVYNLGGNGAMFW 350
 QY 373 MLAGNVNQP-----YPNYGCFNYYPSSTATVLAASE-ALAIATG 411
 DB 351 MLAGIGEGWDRKGYYPDYDGFRIVNDSESEAKLIREYAKILFSTG 396

RESULT 2
 AAW34564
 ID AAW34564 standard; Protein; 680 AA.
 XX
 AC AAW34564;
 XX
 DT 12-MAR-1998 (first entry)
 XX
 DE *Thermotoga maritima* beta-mannanase.
 XX
 KW Glycosidase; thermostable; textile; food processing; pharmaceutical;
 KW detergent; baking; industry; *Thermococcus*; *Staphylothermus*;
 KW *Pyrococcus*; glucose; soluble oligosaccharide; beta-mannanase.
 XX
 OS *Thermotoga maritima*.
 XX
 PN WO9725417-A1.
 XX
 PD 17-JUL-1997.
 XX
 PF 10-JAN-1997; 97WO-US00092.

XX 13-SEP-1996; 96US-0712612.
 PR 11-JAN-1996; 96US-0583787.
 XX
 PA (RECO-) RECOMBINANT BIOCATALYSIS INC.
 XX
 PI Bylina EJ, Lam DE, Mathur EJ, Swanson KV;
 XX
 DR WPI: 1997-372858/34.
 DR N-PSDB: AAT93688.
 XX
 PT New thermostable glycosidase(s) - from *Thermococcus*, *Staphylothermus*
 PT and *Pyrococcus*, used in the textile, food processing,
 PT pharmaceutical, detergent and baking industries
 XX
 PS Claim 4: Fig 11; 82pp; English.
 XX

The present sequence represents beta-mannanase isolated from *Thermotoga maritima*. The enzyme or its encoding nucleic acid sequence is used for generating glucose from soluble oligosaccharides. The enzyme can be used in the food processing, pharmaceutical, textile, detergent and baking industries. The enzyme is also used to treat lactose intolerance, as a diagnostic reporter molecule, in corn wet milling or in the fruit juice industry. The enzymes can be used to hydrolyse gum to remove non-reducing terminal mannose residues. The nucleic acids encoding the enzyme may be used to generate probes to identify similar sequences.

XX
 SO Sequence 680 AA:

Query Match 19.7%; Score 813; DB 18; Length 680;
 Best Local Similarity 41.2%; Pred. No. 7.9e-39;
 Matches 172; Conservative 66; Mismatches 135; Indels 44; Gaps 14;

QY 19 AVAAATLGSITMPSATAAPAGVVTASGQFVNLGLPYRYGGTNNYYLSYQSHADVDDVL 78
 DB 12 SVSAEPLLLIVELSSVLFASDEFVKVENGKFAKNGKFRFGSNYYMYHYKSNRMIDSVL 71
 QY 79 AKAAANLSVIRTWGFIDIGSLDGSVPTIDGNKNGFYFOYWDPSGCAPAYNDG---PTG 134
 DB 72 DSRDNGIKVLRHWG-----LEGSYCRDKN-----TYMHPAGVFGVPEGISNAQSG 120
 QY 135 LOGLDYATASAAHGLRVIVLTNDMKFEGMDQYDKWGLPYHDNFTDPTQOAYKNW 194
 DB 121 FERLDYTVAKAKELGKLVVLNNWDFGGMQYVRFVFCGTHDDFYRDKDKIKEYKKY 180
 QY 195 VNHLLNRVNSITGVTYKNDPTIFAWELANPRCVSGTCLPTSGTCTQATIVNWDQMSAY 254
 DB 181 VSFLVNHVNTYTCVPYREPTIFAWELANPRC-----ETDKSGN-----TLVWEVEMSSY 232
 QY 255 VKSIDPNHMYSGDEGF---YIGSTCGS---WPYNDPSGVDNNALLRVKNDPFTGYHLL 308
 DB 233 IKSLDPNHLVAVDEGFFSNYEGFRPYGGAWAYNG-SSGVDWKLLLEIETVDFGTFHLL 291
 QY 309 YPNYNGON---ADMGTQWIKDHIAAAAIKPTILKEFG--WQTP-DRDSVYQVTTQV 361
 DB 292 YPSHWGVSPEVNAQWGAKWIEDIIKIAKEGKPVVLEEYGIKPSAPVNRVATYRLNDLV 351
 QY 362 RTNGEAGNFWMLAGNVNQP-----YPNYGCFNYYPSSTATVLAASE-ALAIATG 411
 DB 352 YDLGGDGAEMFWMLAGIGESDRDERGYYPDYDGFRIVNDSDSEAEKLIREVAKLFNTG 408

RESULT 3
 AAW35005
 ID AAW35005 standard; Protein; 680 AA.
 XX
 AC AAW35005;
 XX
 DT 21-MAY-1998 (first entry)
 XX
 DE *Thermotoga maritima* endoglucanase.
 XX

Wed Nov 13 17:38:04 2002

KW Endoglucanase; cellulase; carboxymethylcellulose; cellulose;
KW biomass; beta-1,4-glycosidic bond; hydrolysis; saccharification;
KW thermostable enzyme; thermophilic; glycosidase.

OS Thermotoga maritima strain MSB8 (Clone 6GP2).

XX WO9744361-A1.

XX PD 27-NOV-1997.

XX PF 22-MAY-1997; 97WO-US08793.

XX PH 22-MAY-1996; 96US-0651572.

XX PA (KCCO-) RECOMBINANT BIOCATALYSTS INC.

XX PI Lam DE, Mathur EJ;

XX DR WPI: 1998-018435/02.

XX DR N-PSDB: AAT94213.

XX PT Endoglucanase(s), preferably from archaeal bacterium, AEP111a -
PT useful to degrade carboxymethylcellulose and hydrolyse of
PT beta-1,4-glycosidic bonds in cellulose

XX PS Claim 1; Fig 10; 164pp; English.

XX CC This protein comprises an endoglucanase of Thermotoga maritima MSB8
CC (clone 6GP2) that is capable of degrading carboxymethylcellulose
CC and of hydrolysing the beta-1,4-glycosidic bonds in cellulose. It
CC has homology to an endoglucanase of archaeobacterium AEP111a (see
CC AAM34985). It can be produced from native cells or from recombinant
CC host cells, especially prokaryotic host cells transformed with a
CC plasmid or virus-derived vector including the endoglucanase DNA
CC (see AAT94213). 24 Endoglucanases (see AAM34986-W35008) are claimed.
CC They can be used to degrade cellulose for the conversion of plant
CC biomass into fuels and chemicals, for use in detergents, textiles,
CC animal feed, waste treatment, and in the fruit juice and brewing
CC industries for the clarification and extraction of juices.

XX SQ Sequence 680 AA;

Query Match 19.7%; Score 813; DB 19; Length 680;

Best Local Similarity 41.5%; Pred. No. 7.9e-39;

Matches 173; Conservative 65; Mismatches 135; Indels 44; Gaps 13;

OY 19 AVAAATIGSTIMPSATAPACFTASGQFVLNCLPYRYGCTNNYXLSYOSHADVDDVL 78

DB 12 SVSAEFLLLIVELSVLFASDEFYKVENKREKALNCKEERFICSNYYMHYKSNMGIDSVL 71

OY 79 AKAAAMNLSVITKWCFTDGLSGDSVPTIDGNKNGFYQYWDPTGAPAYNDG----PTG 134

DB 72 ESARDMGKLVRLWGF-----LDGESYCRDN-----TYMHPEPGVFGVPEGISNAQSG 120

OY 135 LQGLDYATASAAHGLRVIVVLTNDWKEFGMDQYDKWYGLPYHDNFYDPTQOAYKNW 194

DB 121 PERLDYTVAKAKELGKLVIVLVNWDGFGMNQYVRMFGGTHDDFYRDEKIEEKY 180

OY 195 VNHLLNRVNSITGVYTKNDPTIFAMELANEPCVSGTLPTSGTCTOATIVNWDMSAY 254

DB 181 VSLFNHVNVTYGVYRREPTIMAMELANEPC-----ETDKSGN-----TLVKEVKEMSY 232

OY 255 VKSIDPNHMSVDEGCF---YIGSTQSG---WPYNDPSGDVNDNALLRVKNIIDFGTYHL 308

DB 233 IKSLDPNHLVAVGDEGCFYSNKGFPYCGEAEWAYNGWS-GVDKLLLSITVDFCTEHL 291

OY 309 YPNYWGQN-----ADNGTOWIKDHIAAAIGKPTLIEERG--WQTP-DRDSVYQTWTQTV 361

DB 292 YPSHWGVSPEYAOAGAKWIEDHIIKIAEIGKPVVLEEYGIKPSAPVNRNTALIRLWMDLV 351

OY 362 RTNGEAGNFWMLAGNVNCP-----YPNYDGFNYYHSSTATVLASE-ALAIITG 411

DB 352 YDLGGDGAMFWMLAGIGESDRDERGYYPDYDGFRIVNDSDSPEAEILIREYAKLFNTG 408

RESULT 4

AAW49872

ID AAW49872 standard; Protein: 663 AA.

XX AC AAW49872;

XX DT 21-DEC-1998 (first entry)

XX DE Thermotoga maritima MSB8-6GP2 glycosidase.

XX KW Glycosidase; MSB8-6GP2; thermostable enzyme; oligosaccharide;

XX KW glucose; sugar; baking; textile; detergent; beta-galactosidase.

XX OS Thermotoga maritima strain MSB8-6GP2.

XX PN WO9824799-A1.

XX PD 11-JUN-1998.

XX PF 08-DEC-1997; 97WO-US22623.

XX PR 10-OCT-1997; 97US-0949026.

XX PR 06-DEC-1996; 94US-0056916.

XX PA (DIVE-) DIVERSA CORP.

XX PI Bylina EJ, Lam DE, Mathur EJ, Swanson RV;

XX DR WPI: 1998-362407/31.

XX DR N-PSDB: AAV36921.

XX CC Glycosidase enzymes from organisms of the genera Staphylothermus,
XX CC Pyrococcus and Thermococcus - for deriving sugar from
XX CC oligosaccharides, useful in the e.g. food processing, textile or
XX CC baking industries

XX PS Claim 1; Fig 15a-d; 92pp; English.

XX CC This is the amino acid sequence of glycosidase MSB8-6GP2, deduced
XX CC from a polynucleotide (see AAV36921) of clone 6GP2 of Thermotoga
XX CC maritima MSB8 that grows optimally at 85 degC in high salt medium.
XX CC The invention provides 18 polynucleotides (see AAV36907-24) coding
XX CC for thermostable glycosidases (see AAW49858-75) having glucosidase,
XX CC alpha-galactosidase, beta-galactosidase, beta-mannosidase,
XX CC beta-mannanase, endoglucanase or pullulanase activity. Vectors and
XX CC host cells are also claimed. A method is provided for producing
XX CC the enzymes by recombinant techniques. A claimed method for
XX CC generating glucose from soluble cell oligosaccharides comprises
XX CC contacting a sample (selected from dairy products, fruit juices,
XX CC detergent, textile, guar gum, animal feed, plant biomass or waste
XX CC product) containing oligosaccharides (selected from maltose,
XX CC cellobiose, lactose, sucrose, raffinose, stachyose, verbascoside,
XX CC cellulose, starch, amylose, glycogen, disaccharides, polysaccharides
XX CC and pullulan) with one of the claimed glycosidases such that glucose
XX CC is produced.

XX SQ Sequence 663 AA;

Query Match 19.7%; Score 812; DB 19; Length 663;

Best Local Similarity 42.8%; Pred. No. 8.8e-39;

Matches 169; Conservative 61; Mismatches 121; Indels 44; Gaps 13;

OY 41 FYTASGQFVLNGLPYRYGCTNNYXLSYOSHADVDDVLAKAAAMNLSVITKWCFTDGLSG 100

DB 17 FVKVENKREKALNCKEERFICSNYYMHYKSNMGIDSVLSEARDMGKLVRLWGF-----L 71

OY 101 DGSVPTIDGNKNGFYQYWDPTGAPAYNDG-----PTGLOGLDYATASAAHGLRVIVVL 156

DB 72 DGESYCRDN-----TYMHPEPGVFGVPEGISNAQSGFERLDYTVAKAKELGKLVIVL 125

OY 157 TNDWKEFGMDQYDKWYGLPYHDNFYDPTQOAYKNHLLNRVNSITGVYTKNDPTI 216

116 126 VNWDDFGGNNQYVYKWFGGTHDDFYRDEKIKKEYKKYVSFLVNHVNTYTCVYKREPT 185
117 217 FAWELANEPRCVGSGTLPTSGTCTOATIVWVDDMSAYVKSIDPNHNVSCDECF---YI 273
118 186 MAWELANEPRC-----FTDKSGN-----TLVEWVKEMSSYIKSLDPNHLVAVGDEGFFSNYE 237
119 274 GSTGCGSG---WPNYDPSGDGVNNDALLVKNIIDFCTYHLYPNWQON---ADMCTQWIKI 326
120 238 GFKPYGGEAEAWAYNGWS-GVDMKKLLSTETVDFCTFHLTPSHWGSVPENYAWGAKKIED 296
121 127 HIANAAAIKPKTILEEFG--WOTP-DRDSVYQVWTFQVTRTNGEAGWNFMWLAGVNWGCP- 382
122 297 HIKTAKEIKPVLVEEYGIKPSAPVNTATYRIWDLVYDILGGDCAMFWMLAGICEGSDR 356
123 483 -----YPNYDGFNVYSPSSTATVLASE-ALAIATG 411
124 357 DSRGYYPDYDGFRIVNDPSPEAEILREYAKLFNTG 391

RESULT 5
AAW49868
ID AAW49868 standard; Protein: 680 AA.
XX
AC AAW49868;
XX
XX 21-DEC-1998 (first entry)
DE Thermotoga maritima 6CP2 beta-mannanase.
XX
KW Glycosidase; 6CP2; thermostable enzyme; oligosaccharide; glucose;
KW sugar; baking; textile; detergent; beta-mannanase.
XX
OS Thermotoga maritima strain 6CP2.
IN W09824799-A1.
XX
XX 11-JUN-1998.
XX
XX 08-DEC-1997; 97MO-US22623.
XX
XX 10-OCT-1997; 97US-0949026.
XX 06-DEC-1996; 96US-0056916.
XX
XX (DIVE-) DIVERSA CORP.
XX
XX Bylina EJ, Lam DE, Mathur EJ, Swanson RV;
PI
XX
XX WPI; 1998-362407/31.
XX N-PSDB; AAV36917.
XX
XX Glycosidase enzymes from organisms of the genera Staphylothermus,
PT Pyrococcus and Thermococcus - for deriving sugar "n"
PT oligosaccharides, useful in the e.g. food processing, textile or
PT baking industries
XX
XX Claim 1; Fig 11a-d; 92pp; English.
XX
XX This is the amino acid sequence of beta-mannanase 6CP2, deduced
CC from an isolated polynucleotide (see AAV36917) of a Thermotoga maritima
CC clone (6CP2) that grows optimally at 85 degC in high salt medium.
CC The sequence shows 56% amino acid identity to a mannanase of
CC Aspergillus aculeatus. The invention provides 18 polynucleotides
CC (see AAV36917-24) coding for thermostable glycosidases (see AAW49858-75)
CC having glucosidase, alpha-galactosidase, beta-galactosidase,
CC beta-mannosidase, beta-mannanase, endoglucanase or pullulanase
CC activity. Vectors and host cells are also claimed. A method is
CC provided for producing the enzymes by recombinant techniques. A
CC claim is made for generating glucose from soluble cell
CC oligosaccharides comprises contacting a sample (selected from dairy
CC products, fruit juice, detergent, textile, guar gum, animal feed,
CC plant biomass or waste product) containing oligosaccharides
CC (selected from maltose, cellobiose, lactose, sucrose, raffinose,

CC stachyose, verbascose, cellulose, starch, amylose, glycogen,
CC disaccharides, polysaccharides and pullulan) with one of the
CC claimed glycosidases such that glucose is produced.
XX
SQ Sequence 680 AA;
Query Match 19.6%; Score 806; Dk 19; Length 680;
Best Local Similarity 41.2%; Pred. No. 2c-18;
Matches 172; Conservative 65; Mismatches 136; Indels 44; Gaps 14;
QY 19 AVAAATLGSITMPSATAAPAGFVTASGQFVLNCLPYKYGTNNYLYSYQSHAIWUWIL 74
Db 12 SVSAEFLLLIVELSFVLPASDEFVKVENGKFKALNGKEKFKIGSNYYMHYKSNGLDSVL 71
QY 79 AKQAMNLSVIRTWGFTIDIGSLDGSVPTIDGNKNGFYFOYWDPSGAPAYNDG-----PTG 134
Db 72 ESARDMGIKVLRIWGF-----LDGESYCRDN-----TYMHPEPGVFGVEGTSNAUGS 120
QY 135 LOGLDYATASAAHCLRVIVLTNPWKKEFGCMDOYDKWYGLPYHNFYTDHPKQAYKNW 194
Db 121 FERLDYTVAKAKELGKLVIVLVNMDDFGGMNVQVVRWPGGTHHDDPYRDEKIKKEYKKY 180
QY 195 VNHLLNRVNSITGVTYKNDPTIFAMELANEPRCVGSGTLPTSGTCTQATIVWVDDMSAY 254
Db 181 VSFLVNHVNTYTCVYKREPTIMAMELANEPRC-----FTDKSGN-----TLVEWVKEMSSY 242
QY 255 VKSIDPNHNVSGDEGF---YIGSTGSG---WPNYDPSGDGVNNDALLVKNIIDFCTYHIL 408
Db 233 TKSLDPNHLVAVGDEGFFSNYEGFKPYGGEAEAWAYNGWS-GVDMKKLLSTETVDFCTPHL 291
QY 309 YPNYWGQN---ADMCTQWIKDHIANAAAIKPKTILEEFG--WOTP-DRDSVYQVWTFQV 461
Db 292 YPSHWGVSPENYAONGAKWIEDHIIKIAKEIGKPVVLEEYGIKPSAPVNTATYRIWDLV 351
QY 362 RTNGEAGWNFMWLAGVNWGCP-----YPNYDGFNVYSPSSTATVLASE-ALAIATG 411
Db 352 YDLGGDCAMFWMLAGICEGSDRDEKCYYPDYDGFRIVNDPSPEAEILREYAKLFNTG 408

RESULT 6
AAV69508
ID AAV69508 standard; Protein: 562 AA.
XX
AC AAV69508;
XX
XX 10-APR-2000 (first entry)
XX
XX Acidothermus cellulolyticus E1 endoglucanase.
DE
XX
KW E1 endoglucanase; cellulose binding domain; CHD; cellulose modification;
KW beta-1,4-endoglucanase; endocellulase; thermostable.
XX
OS Acidothermus cellulolyticus.
XX
XX Key Location/Qualifiers
XX Peptide 1..41
XX Peptide /note= "Putative signal peptide"
XX Peptide 14..41
XX Peptide /note= "Putative signal peptide (alternative)"
XX Domain 42..404
XX Domain /note= "Catalytic domain"
XX Region 405..460
XX Region /note= "Linker region"
XX Domain 461..562
XX Domain /note= "Cellulose binding domain (CHD)"
XX
XX CA2226898-A1.
PN
XX
XX 25-SEP-1999.
PD
XX
XX 25-MAR-1998; 98CA-2226898.
PF
XX
XX 25-MAR-1998; 98CA-2226898.
PR

[illegible]

RESULT 9	
AAK89927	
ID	AAK89927 standard; Protein: 521 AA.
XX	
XX	AAK89927:
XX	
XX	08-OCT-1996 (first entry)
XX	
XX	A. cellulolyticus E1 endoglucanase.
XX	
XX	E1 endoglucanase; cellulase; cellulose; saccharification; ethanol.
KW	
XX	Acidothermus cellulolyticus.
OS	
XX	
PH	Key Location/Qualifiers
FT	Active-site
FT	Domain
FT	1..363
FT	/label= Catalytic-domain
FT	364..417
FT	/label= Linker
FT	/note= "proline/serine/threonine-rich linker
FT	domain common to multi-domain microbial
FT	cellulases"
FT	

[illegible]

[illegible]

XX DE Acidothermus cellulolyticus E1 endoglucanase mutant, Y245G.
XX KW E1 endoglucanase; glycosyl hydrolase; insoluble substrate;
XX KW cellulose hydrolysis; ethanol production; fermentation;
XX KW mutant; mutein.
XX XX
XX OS Acidothermus cellulolyticus.
XX OS Synthetic.
XX PN WO200070031-A1.
XX PD 23-NOV-2000.
XX XX
XX XX 19-MAY-2000; 2000WO-US13971.
XX XX 19-MAY-1999; 99US-0134925.
XX XX (MIDE) MIDWEST RES INST.
XX PA
XX XX Himmel ME, Adney WS, Baker JO, Vinzant TB, Thomas SR, Sakon J;
XX P1 Decker SR;
XX P1
XX XX WPI; 2001-061226/07.
XX DR
XX XX Preparation of glycosyl hydrolase with an increased catalytic activity
XX PT on insoluble substrate
XX PT
XX XX
XX PS Claim 5; Page 27-29; 30pp; English.
XX XX
XX CC The invention relates to a method for making glycosyl hydrolase mutants
XX CC with increased catalytic activity with either insoluble or soluble
XX CC cellulose substrates relative to the wild-type enzyme. The method for
XX CC making a glycosyl hydrolase with increased insoluble substrate catalytic
XX CC activity comprises replacing an active site-associated amino glycosyl-
XX CC stabilising amino acid with an amino acid that does not strongly bind a
XX CC disaccharide product in the active site. Conversely, the method for
XX CC making a glycosyl hydrolase with increased soluble substrate catalytic
XX CC activity comprises replacing a hydrophobic substrate-binding amino acid
XX CC with a positively charged residue. The invention also discloses mutants
XX CC of Acidothermus cellulolyticus E1 endoglucanase (AAB48786-B48788)
XX CC produced according to the method of the invention. The Y245G mutant
XX CC (AAB48788) has improved activity with insoluble substrates, and the W42R
XX CC (AAB48786) and Y82R (AAB48787) mutants have improved activity with
XX CC soluble substrates. The invention also encompasses DNA encoding these
XX CC mutants. The glycosyl hydrolases of the invention are used as catalysts
XX CC for cellulose hydrolysis to produce sugars that can be fermented to
XX CC produce fuels such as ethanol. The present sequence represents the
XX CC Acidothermus cellulolyticus E1 endoglucanase Y245G mutant.
XX CC
XX S0

Query Match 16.2%; Score 668.5; DB 22; Length 521;
Best Local Similarity 27.1%; Pred. No. 1.2e-30;
Matches 210; Conservative 87; Mismatches 176; Indels 301; Gaps 31;

QY 37 APAGFVTASGQPV-LNGLPYRYGGTN-----NYLSYQSHADVDVLAQAAMLSV 88
Db 1 AGCGYHRTSGREILDANNVPRVIRAGNFGFETCNVWHGLWSRDYRSLDQIKSLGYNT 60
QY 89 IRTWGFIDIGSLDGSVPTIDGNKNGFYFQYWDPTGAPAYNDGPTGLQGLDYATASAAH 148
Db 61 IRL-PYSDDILKPTMP-----NSINFYQMNQD-----LOGLTSLQVMDKIVAVAGQI 107
QY 149 GLKRVIVLVNMDKFEFGMDQYDKWGLPYHDNFYDPTQQAYKNVNHLLNRVNSITGV 208
Db 108 GLRII--LDHRHPCDSCG--QSALWY-----TSSVSEATWISDL-----QALAQ 146
QY 209 TYKNDPTIFAWELANPR---CVSGSTLPTSGTCTQATVNW---VDQMSAYVKSIDPNH 262
Db 147 RYKGNPTVVGFDLRNPHDPACGCCD-PS-----IDWRLAAERAGNAGNVLVSNPNL 196
QY 263 MVSV-----GDEGYIGSTQSG-WPYNDPSGDGVNALLRVKNIDFGTYHLY----- 309

Db 197 LIFVEGVQSYNGDSYWMGNGNLQAGQYP-----VVLNVPNRLVYSAHDYATSVG 245
QY 310 -----PNYWGQNADMGCTOWIKDHIANAAIAIGKPTLLEEGFMQTPDRDSYVQ 355
Db 246 PQTWFSDPPTFPNNMPGIWNKN--WG-YLFNQNJA-----PVMLGEGF--TLQSTTDQ 293
QY 356 TWTOTVRTNGEAGNFWMLAGNVNGOPYPNYDGFNVYPSSTATVLAASEALAIATGTSHP 415
Db 294 TWLKTLL----- 299
QY 416 PPSSSSPSSP 475
Db 300 -----VQY-----L 304
QY 476 KPGQLVNTGSSVDLSTVTVVRYWFTRDGSSSTLYVNCDMAAMCGNIRASFGSVNPAIP 535
Db 304 RP-----TAQY-----CADSFQWTFWSNPNPS- 325
QY 536 TADTYLQLSFTGGTLAAGSGTGEIONRVNKSMDSNFDETDNDYSYCTNTAFQDWTKVTVV 595
Db 326 -----GDTGGIL-----KDDMTVDTVKD-----GYLAPIKSSIP- 355
QY 596 NGRLVMGTEPSTPSTPSTPSTPSTPSTPSTPSTPSTPSTPSTPSTPSTPSTPSTP 649
Db 356 -----DPVGASAPSSQSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSP 407
QY 650 PPSPPSPS-VSSSGVCGRATYVNVNSDMGSGTATVTVTNTGSRATSGMTVAMSGFNVIV 70H
Db 408 PTASPTSPSTAASCARCTASYQVNSDMGNGFTVAVVNTSGSVATKTVTSWTEGNGOTJ 467
QY 709 TNYWNTALTQSGASVYATNLNNVNIQPCQSTTFGFCNGSYSGTNTAPTTLTCTAS 762
Db 468 TNSWNAAVTONGQSVTARNNSYNNVIQPCQNTTFGFOASYTGSNAAPTVAACAAS 521

RESULT 12
AAY13493
ID AAY13493 standard; Protein; 1751 AA.
XX AC AAY13493;
XX DT 30-JUL-1999 (first entry)
XX DE Truncated cellulases comprising amino acid sequence.
XX KW Cellulase; proteinase; truncated; Cel B5; Cel B4/5; Cel E1; Cel E1/2;
XX KW Cel 1/2/3; Cel 6; Cel E3/B5; laundry detergent; stain-removing;
XX KW cotton-containing fabric; stonewashing.
XX OS Unidentified.
XX PN EP921188-A2.
XX PD 09-JUN-1999.
XX PF 15-SEP-1998; 98EP-0810919.
XX PR 19-SEP-1997; 97US-0932571.
XX PA (CLRN) CLARIANT FINANCE BV1 LTD.
XX PI Anderson P, Bergquist PL, Daniels KM, Farrington GK;
XX PI Gibbs MD, Morgan H, Williams DP;
XX DR N-PSDB; AAX55662.
XX PT WPI; 1999-315403/27.
XX PS New truncated cellulase proteins, useful in detergents and for
XX PS producing 'stonewashed' denim
XX PS Claim 7; Page 37-41; 65pp; English.

The invention relates to a recombinant cellulase active protein free of proteinases of native thermophilic and alkaliphilic origin, comprising the truncated sequences Cel B5, Cel B4/5, Cel E1, Cel E1/2, Cel 1/2/3, Cel 6 or Cel E3/B5, or a stability region from one of the defined full-length sequences, or functional equivalents. Cel B5 extends from amino acid A1011 to P1424 or K1425 or N1426, and Cel B4/5 extends from amino acid K635 to N1426 in the sequence shown in AA13492; Cel E1 extends from amino acid Y39 to D481, Cel E1/2 extends from Y39 to G635, Cel E1/2/3 extends from Y39 to G812, Cel E6 extends from amino acid V123 to K1751 and the stability region extends from amino acid E482 to G635 in the sequence shown in AA13493; Cel E3/B5 is shown in AA13494. The new enzymes are useful in laundry detergent compositions to prevent or remove staining, backstaining or graying, for use on cellulosic materials including cotton-containing fabrics. They are especially useful for preventing redeposition of colorant during stonewashing, and for processing of textiles where cellulose breakdown is required. The new truncated enzymes show reduced redeposition of dye compared to using non-truncated cellulase compositions.

Sequence	1751 AA;
5Q	

Query Match 15.1%; Score 624; DB 20; Length 1751;
Best Local Similarity 25.6%; Pred. No. 1.8e-27;
Matches 234; Conservative 140; Mismatches 278; Indels 262

Qy	7	PARAFVA-TAAGTAVAAANATLGSITWPSA-----TAAPAGFVTASG-----	46
Db	183	PASSVVAETAAASLAVASVJIKERNOSKAASYLOHAKDLFEFADTTKSDAGCYTAATGYTS	242
Qy	47	GQFVLN---GLPYRYGCTNNYILSYOSHADVDDVLAKAQAAMNLSVTRWCFIDIGSL---	100
Db	243	GGFIDDLGMAAVWLYIATND--SSYLTKA--EELMEYANGTWTWQCHDDVRYGTLIML	298
Qy	101	-----DCSVPTIDCNKNGHYFYQW-DPSTCAPAYNDCPTGICGLDYATASAAHGL	150
Db	299	AKITGKELYKGAV-----ERN-----LDHWTDRITYTPKGMAYLTGWSGLRYATTA--	345
Qy	151	KVIVLVLTNDNKEFGGMDYDKW-----YGLPYHDNFYDPRTOQA	190
Db	346	-FIACVYADW---SGCISNKKTKYLNPAKSOIDYALGSGTCRSPVVGCGTNYPOHPHIRNA	401
Qy	191	YKNWVW-----ILLIRVNSITG-----VTYKNDPTIFANELANERPC-VGSGTLPT	235
Db	402	ISSWANSMKIPEYIIRHJL--YGALVGCGPGSDSYNDIDTY---VQNEVACDYNAGIVGA	456
Qy	236	SGTCTQATIVNWVDMSAYVKSIDPNHWSVGDEGF---IGSTOG-----S	279
Db	457	LAKWYLYGCEPIDDF-----KAIE-----TPTNDEIFVESKFGNSGPNTEVILSY	508
Qy	280	GWP-----YNDPSPGVDDNALLRVKNIDFGCTIILY-----PNYWQGNAD---	318
Db	509	GWP-PRVTDKLSFKYFIDITELIOAGYSPPDVVKVD--TYVIEGCKISGPYWDKNRNIYV	566
Qy	319	---WGTOWIK---DIIIAANAIGXPTILEEFECWOTPDORSVYQWTQTVRTNGEAGWN	370
Db	567	LVDFSGTKIYPGGEVEIHKQADF--KLSPVQGYPWD-PTNDPSPYKGLTSOLEKN-----	617
Qy	371	FWMLAGNVNCOOPYNDYDGFNVYYPSSTATVLAASEALAISTGTPPPSPSSSPSSPSP	430
Db	618	-----KYIAAYDNNNLVW-----GLEPGAATSTP-----APTSTPTTP	651
Qy	431	SPSPASPSASSPSSPSSSPSSPVSG-GVKVOYKKNDSAPDNQIKPGLQVLWNTGSSSV	489
Db	652	TPITPTVATP--TPITPTPTPGSGGVKVLVKNNETSASTGSLRPWFKIVNGCGSSV	709
Qy	490	DLSTVTRYWFTRDGGSSTLVYKCDWAAMCGGNIRASFGSVNPAITPATDYTLQLSFTGCT	549
Db	710	DLSRVKIRYWTVDGDKPOSAB--CDWAQICASNVTENFVKLSGVSYADYILEVGFSSA	768
Qy	550	--LAACGSFCEIQNRNVKNSDWSNFDETNDYS-----YGTNTAFODWTKVTYVNGRLV	600
Db	769	GQLOPKDQTDIQVRFNKNDWNSYNQADMSLOSMTNYGEN-----AKVLYLYVDGVLV	822

CC The present sequence is Cele protein.

```
XX SQ Sequence 1751 AA;
Query Match 15.1%; Score 624; DB 23; Length 1751;
Best Local Similarity 25.6%; Pred. No. 1.4e-27;
Matches 234; Conservative 140; Mismatches 278; Indels 262; Gaps 46;

QY 7 PARAFVA--TAAGTAVAAATLGSITMPA-----TAAPAGFTVTSAG--- 46
DB 183 PASSVVAETAASLAVASVVLKERNQKAASVQLQAKDLFEADTTRSDAGTAAATGYTS 242
QY 47 GQFVLN---GLPVRGTTNNYLYSYOSHADVDDVLAKAAMNLSVIRTWGFDIGSL--- 100
DB 243 GGFIDDLGMAAVWLYIATND--SSYLTKA--EELMSEYANGTNTWTQCDVDRYGTLM 298
QY 101 -----DCSVPTIDGKNGFYFYQW--DPSTCAPAYNDGPTCLQGLDYAIASAAHGL 150
DB 299 AKITGELYKGAV-----ERN-----LDHMTORITVTPKGMAYLTGWSGLRYATTAA----- 345
QY 151 RVIVLTNDMKKEFCGMQDYDKW-----YGLPYHDNFYTDPRTOQA 190
DB 346 -FLACVYADW---SGCDSSNKTKYLNFAKSOIDYALGCTCRSFVVGFTNYPOHPHHRNA 401
QY 191 YKMWVN-----HLLNRVNSITG-----VTYKNDPTIFAMELANEPRC--VCSGTLPT 235
DB 402 HSSWANSMKIPEYHRHIL--YGALVCGPGSDSYNDIDTDY---VQNEVACDYNAGIVGA 456
QY 236 SGCTCTQATIVNWDOKSAYVKSIDPNHIVSVGDEGFY---IGSTQG-----S 279
DB 457 LAKMYOLYGEPIDDF-----KALE-----TPTNDEIFVESKFGNSQSGFTYVSYINRT 508
QY 280 GWP-----YNDPSGDVNNALLRVKNIDFGTYHLY-----PNYWGQAD----- 318
DB 509 GWPFRVTDKLSFYKFDLTLELIQAGYSPDVVKVD--TYTIEGCKISCPYVMDKRNLIYV 566
QY 319 -----WGTQWIK-----DHIANAAIGPRTILEEFCWQTPDRDSVYQVWTQVTRNGEAGWN 370
DB 567 LVDFSTCKIYPGGEVEHKKQAOF--KISVPQGYPWD--PTNDPSYKGLTSOLEKN----- 617
QY 371 FWMLAGNVAGQPNPDGPNVYPSSTATVLAELALISTGTPPPSPSSPSSSPSSPSP 430
DB 618 -----KYTAADNNNLVW-----GLEPGNAATSTP-----APTSTPTPTP 651
QY 431 SPASPSASPSASSPSPSPSPSPSVSG-GVKVQYKNDNSAPGDNQIKPGLQLVNTGSSSV 489
DB 652 TPTPTVTATP--TPTPTPTPTGSPGSGVKVLYKNNETSASTGSIKPFWKIVNGGSSSV 709
QY 490 DLSTVTYRWFTRDGGSTLVYNCWNAACCGNIRASFGSVNPTATPTADTYLQSLFTGGT 549
DB 710 DLSRVKIRIYVTVVDDGKPOSAY-CDWAQICASNVTNFVKLSGSGVSGADYVLEVPSSGA 768
QY 550 --LAAGCGSTGEIQNRVYNKSDWSNFDETNDYS-----YCTNTAFQDWTKVTVVYVNGRLV 600
DB 769 GOLQPGKDTGDIQVRFNKNDWSNYNQADDMSWLMQNTNYGEN-----AKVLYVDGVLV 822
QY 601 WGTERTSPTSPPTSPPTSPPTSPPTSPPTSPPTSPPTSPPTSPPTSPPTSPPTSP 654
DB 823 WQEPFGGATPAPTSTATPTPTATPTPTP--TPTPTPTVSATPTPTAPTASPVGGSYWTP 880
QY 655 SPVSSSSGVGRATVYVNSDWSG---FTATVTVTNGS-----RAT 693
DB 881 SES-----YGALKVYANGNLSSPTNLPKIKIENVGTGTAVDLSRVKRYVYTTIDGEAT 935
QY 694 SCWTVAMS-----FCGNTQVTNYMNTALTOSCATVATNLSNNV---LQPGQ 738
DB 936 QSVSVASSINPAYIDVKFVKLCANAGGADYYEIGFKSGAGVLACQSTKEIRLSQK- 994
QY 739 STTFGFGNSYSGTN 752
DB 995 -----SGSYNQSN 1002
```

RESULT 14

AAV13492
ID AAY13492 standard; Protein; 1426 AA.

XX AC AAY13492;

DT 30-JUL-1999 (first entry)

XX DE Truncated cellulases Cel B4/5 and Cel B5.

XX KW Cellulase; proteinase; truncated; Cel B5; Cel B4/5; Cel E1; Cel E1/2;
KW Cel 1/2/3; Cel 6; Cel E3/B5; laundry detergent; stain-removing;
KW cotton-containing fabric; stonewashing.

XX OS Unidentified.

XX PN EP921188-A2.

XX PD 09-JUN-1999.

XX PF 15-SEP-1998; 98EP-0810919.

XX PR 19-SEP-1997; 97US-0932571.

XX PA (CLRN) CLARIANT FINANCE BVI LTD.

XX PI Anderson P, Bergquist PL, Daniels RM, Farrington GK;
PI Gibbs MD, Morgan H, Williams DP;

XX DR WP1; 1999-315403/27.

XX DR N-PSDB; AAX55661.

XX PT New truncated cellulase proteins, useful in detergents and for
PT producing 'stonewashed' denim

XX PS Claim 7; Page 34-37; 65pp; English.

XX CC The invention relates to a recombinant cellulase active protein free of
CC proteinases of native thermophilic and alkaliphilic origin, comprising
CC the truncated sequences Cel B5, Cel B4/5, Cel E1, Cel E1/2, Cel 1/2/3,
CC Cel 6 or Cel E3/B5, or a stability region from one of the defined full-
CC length sequences, or functional equivalents. Cel B5 extends from amino
CC acid A1011 to P1424 or N1426, and Cel B4/5 extends from amino
CC acid K635 to N1426 in the sequence shown in AAY13492; Cel E1 extends
CC from amino acid Y39 to D481; Cel E1/2 extends from Y39 to G635; Cel
CC E1/2/3 extends from Y39 to G812; Cel E6 extends from amino acid V1233 to
CC K1751 and the stability region extends from amino acid E482 to G635 in
CC the sequence shown in AAY13493; Cel E3/B5 is shown in AAY13494. The new
CC enzymes are useful in laundry detergent compositions to prevent or
CC remove staining, backstaining or graying, for use on cellulosic
CC materials including cotton-containing fabrics. They are especially useful
CC for preventing redeposition of colorant during stonewashing, and for
CC processing of textiles where cellulose breakdown is required. The new
CC truncated enzymes show reduced redeposition of dye compared to using
CC non-truncated cellulase compositions.

XX SQ Sequence 1426 AA;

Query Match 14.3%; Score 589.5; DB 20; Length 1426;
Best Local Similarity 24.6%; Pred. No. 1.4e-25;
Matches 199; Conservative 138; Mismatches 251; Indels 221; Gaps 33;

QY 84 MNLISVIRTWCFDI---GSLDGSVPPTI-DGNKNGFYFOYWDPS-----" 123

DB 19 MSISLISGMSYFPVKTEAAPDWSIPSLWESYKNDFKIGVAIPARCLNSDNDKQMLKHN 78

QY 124 GAPAYND-----GPTGLQGLDYAIASAAAH-----GLRVIVL-----TNDWKE 162

DB 79 SITAENKMPESLLAGOTS-TGLSYRFSTADTFTVNTANTNIGTRHTLVVHHNOTPDW-- 135

QY 163 FCGMDQDYKWGLPYHDNFYTDPTQOAYKNWYNHLLNRVNS-ITGVYTKNDPTIFAWEL 221

DB 136 -----FFRDSGGQMLSK---DALLARLKQVIYDVGVGRYKCKVYADV 174

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 13, 2002, 11:50:41 : Search time 35.5819 Seconds
(without alignments)
2058.756 Million cell updates/sec

Title: US-09-917-378-1
Perfect score: 4122
Sequence: 1 MGLVRRPARAFVATAAGTAV.....GFNGSYSGTNTAPTLCTAS 762

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues 283224
Total number of hits satisfying chosen parameters:

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_73:.*
1: pir1:.*
2: pir2:.*
3: pir3:.*
4: pir4:.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	811	19.7	669	D72278	endo-1,4-beta-mann
2	676.5	16.4	1331	A48954	mannan endo-1,4-be
3	661.5	16.0	1711	T31337	1,4-beta-glucanase
4	632.5	15.3	1742	T17120	cellulase (EC 3.2.
5	609	14.8	1779	T31085	xylanase - Caldice
6	603.5	14.6	1039	S02711	cellulase (EC 3.2.
7	526.5	12.8	411	D86153	hypothetical prote
8	498	12.1	448	T48214	endo-1,4-beta-mann
9	495	12.0	431	T09048	probable mannan en
10	484	11.7	1854	S36859	c1pA protein - Clo
11	460	11.2	508	G69593	cellulase (EC 3.2.
12	454	11.0	508	A26874	cellulase (EC 3.2.
13	451.5	11.0	499	A27198	cellulase (EC 3.2.
14	451	10.9	504	S54744	cellulase (EC 3.2.
15	451	10.9	747	B47093	cellulase (EC 3.2.
16	447	10.8	499	JN0111	cellulase (EC 3.2.
17	445	10.8	700	B41897	cellulase (EC 3.2.
18	426.5	10.3	369	T04323	mannan endo-1,4-be
19	426.5	10.3	505	S39962	endoglucanase - Er
20	423.5	10.3	442	F87348	mannanase, probabl
21	415.5	10.1	403	A84592	(1-4)-beta-mannan
22	412.5	10.0	915	A43802	cellulase (EC 3.2.
23	402	9.8	486	T40548	bifunctional cellu
24	394	9.6	145	A41897	cellulase homolog
25	380	9.2	441	T12011	cellulase (EC 3.2.
26	365.5	8.9	890	T35237	probable secreted
27	362	8.8	962	S03818	carboxymethylcellu
28	361.5	8.8	473	S50755	hypothetical prote
29	353.5	8.6	570	S56132	cellulase (EC 3.2.

30	335	8.1	1045	2	A39199	endoglucanase (E
31	330	8.0	872	2	S49541	cellulase - Cellul
32	313.5	7.6	1090	2	S59077	cellulose 1,4-beta
33	312.5	7.6	586	2	PC6006	scaffolding protei
34	311.5	7.6	484	1	A24994	cellulose 1,4-beta
35	311.5	7.6	1162	2	T30433	scaffolding protei
36	310	7.5	611	2	S76211	hypothetical prote
37	308.5	7.5	986	2	SI2021	thermoactive cellu
38	308	7.5	1483	2	C97012	probably celluloso
39	305.5	7.4	1230	2	S47466	cellulose 1,4-beta
40	301.5	7.3	579	2	T35240	probable secreted
41	300	7.3	279	2	T10361	hypothetical prote
42	298	7.2	879	2	A47704	endoglucanase 1 (E
43	294.5	7.1	2232	2	T34434	hypothetical prote
44	285.5	6.9	474	2	SI5921	protein TPX-VT3 -
45	285.5	6.9	1032	2	T34433	hypothetical prote

ALIGNMENTS

RESULT 1
D72278
endo-1,4-beta-mannosidase - Thermotoga maritima (strain MSB8)
C:Species: Thermotoga maritima
C:Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 21-Jul-2000
C:Accession: D72278
R:Nelson, K.E.; Clayton, R.A.; Gill, S.R.; Gwinn, M.L.; Dodson, R.J.; Haft, D.H.; Hic
Garrett, M.M.; Stewart, A.M.; Colton, M.D.; Pratt, M.S.; Phillips, C.A.; Richardson,
C.M.
Nature 399, 323-329, 1999
A:Title: Evidence for lateral gene transfer between Archaea and Bacteria from genome
A:Reference number: A72200; MUID:99287316; PMID:10360571
A:Accession: D72278
A:Molecule type: DNA
A:Status: preliminary
A:Residues: 1-669 <ARN>
A:Cross-references: GB:AE001779; GB:AE000512; NID:q4981777; PIDN:AND36302.1; PID:q498
A:Experimental source: strain MSB8
C:Genetics:
A:Gene: TM1227

Query Match	19.7%	Score	811	DB	2	Length	669
Best Local Similarity	42.8%	Pred. No.	2e+33				
Matches	169	Conservative	61	Mismatches	121	Indels	44
Gaps	13						
Qy	41	FVTASGGQFVLNGLPYRYGGTNNYVLSYQSHADVDVLLAKAQAAMNLSVIRTWGFIIDIGSI	100				
Db	23	FKVKGKFKALNGKEFRFGISNNYMHYKSNRMIDSVLESARDMGIKVLRWGF-----L	77				
Qy	101	DGSVPTIDGNKNGFYFOYMDPSTCAPAYNDG-----PTGLQGLDYAIAASAAHGLRVIIVL	156				
Db	78	DGESYCRDKN-----TYMHPEPGVFGVPEGTSNAQSGFERLDYTVAKAKELGIKLVIVL	131				
Qy	157	TNDKKEGGMDQYDKWGLPYHDNFYDTPRTOQAYKMWVHLLNRVNSITGVTYKNDPTI	216				
Db	132	VNNWDDFGGMNQYVRWFGTHHDDFYRDEKIKEEYKYVSFLVNHVNTYTGVPYREPTI	191				
Qy	217	FAWELANEPKRCVSGSLTPTSGCTCTOATIVNWDQMSAVKSIDPNHMYSVGDEGF----	273				
Db	192	MAWELANEPK-----ETDKSGN-----TLVHWVKEMSSYIKSLDPNHLVAVGDEGFSNYE	243				
Qy	274	GSTQSG-----WPYNDDPGVDNNALLRVKNTDFCTYHLYPNYWGQN-----ADMGQTQIKD	326				
Db	244	GFKPYGGEAWAYNGWS-GVDWKLLSTETVDFCTFHLYPSHWGVSPEYNAQWAKWIED	302				
Qy	327	HIAAAAIGKPTILEERG--WQTP-DRDSVYQTWTQVTRTNGEAGHWFMLAGNVNCGP-	382				
Db	303	HFKIAKEIGKPVVLEEYGIKPSAPVNRATYRLWNLVDYLDLGGDGFAMFMLAGICEGSDR	362				
Qy	383	-----YPNYGDNFYYPYSSTATVLAASE-ALAIISG	411				
Db	363	DERGYYPDYDGFRIVNDSDPEALIREYAKLFNTG	397				

RESULT 2

A48954

mannan endo-1,4-beta-mannosidase (EC 3.2.1.78) - Caldocellum saccharolyticum

N:Alternate names: beta-mannanase

C:Species: Caldocellum saccharolyticum

C:Date: 19-Dec-1993 #sequence_revision 18-Nov-1994 #text_change 08-Oct-1999

C:Accession: A48954; B43745

R:Gibbs, M.D.; Saul, D.J.; Luthi, E.; Bergquist, P.L.

Appl. Environ. Microbiol. 58, 3864-3867, 1992

A:Title: The beta-mannanase from "Caldocellum saccharolyticum" is part of a multidomain

A:Reference number: A48954; MUID:93119139; PMID:1476429

A:Accession: A48954

A>Status: preliminary

A:Molecule type: nucleic acid

A:Residues: 1-1331 <GIB>

A:Cross-references: GB:L01257; NID:g144290; PIDN:AAA71887.1; PID:g144291

A:Note: sequence extracted from NCBI backbone (NCBIN:121576, NCBI:P121577)

R:Luehti, E.; Bhana Jasmat, N.; Grayling, R.A.; Love, D.R.; Bergquist, P.L.

Appl. Environ. Microbiol. 57, 694-700, 1991

A:Title: Cloning, sequence analysis, and expression in Escherichia coli of a gene coding

A:Reference number: A43745; MUID:91247819; PMID:2039230

A:Accession: B43745

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-337, 'PPRQHQRQ' <LUE>

A:Cross-references: EMBL:M36063; NID:g144292; PIDN:AAA72861.1; PID:g144294

A:Note: the authors translated the codon CAC for residue 262 as Glu

A:Note: this sequence has been revised in reference A48954

C:Keywords: glycosidase; hydrolase; polysaccharide degradation

Query Match 16.4%; Score 676.5; DB 2; Length 1331;

Best Local Similarity 24.5%; Pred. No. 2.3e-26;

Matches 208; Conservative 135; Mismatches 262; Indels 243; Gaps 33;

QY 60 GTNNYLYSYQSHADVDVLAQAAMNLSVIRTWGFDIGSLDGSVPITDGNKNGFYQY 119

Db 55 GTNHAHCWYDRDLDT-----ALRGIRSMGM-----NSRVVV--LSNGY---RW 92

QY 120 DPSTGAPAYNDGPTGLGLDYATASAAHGLRVIVLTND-----W 160

Db 93 TKIPASIVAN-----TSLRSISGFKATILEVHDTGYGEGDGAACSLAQAVEYW 141

QY 161 KEFGG-MDOYDKWYGLPYHDNFTDPTQOAYKNWVHLLNRVNSITGVYTKN-----DPT 215

Db 142 KEIKSVLDGNEDEVII---NIGNEPYGNVNNYQWVNDTKNAIKALRDAGFKTKIMVDAP 197

QY 216 IFAWELANEPKRCVSGTLPITSG-----TCTQATIVNVDMSAYVKS-IDPNHMSVGD 268

Db 198 MWGODWSNTRDINAQSIMEADPLRNLFVSIHMYGVYNTASKVEEYIKSFVDKGLPLVIGE 257

QY 269 EGFYIGSTGSGWPYNDPDSGVNNALLRVK-----NIDFGTYHLHPNYWG 314

Db 258 FGHQ--HTDG-----DP-----DEEALVRAKYKICLFSWSWCGNSYVGYLDVNNWD 305

QY 315 QN---ADWGTONIKDHIAAAAIKPT----- 338

Db 306 PNNPTPMG-QWYKTNATIGTSTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPT 364

QY 339 -----ILLEFG----- 359

Db 365 QTKVLYANKETSTNTIRPWLKVVNSGSSIDLSRVTRYWYVDGERAQSAISDWAQ 424

QY 360 TVRTNGEAGNENWMLAGNVGOPY-----PNYDG 388

Db 425 IGASN--VTFKFKVLSVSGADYILEIGFKSGAGQLQPKDGTGEIQMRNFKDDWSNYQ 482

QY 389 FN-----VYPPSATVTLASEALAI-----TGTSPPSPSSSSSSSSSSSSPSASP 436

Db 483 GNDWSIQSMTSYGENEKTAYIDGLVMQGEPSGATPAPAPTATPTPTPTPTPTPTPTPT 542

QY 437 -----SASPSSSSSSPSSSPSVSGG-VKVOYKNNDAPCDNQIKPGLQLVNTGSSVDL 491

Db 543 TPTVTATPTPTPTPTPVSTPATGGQIKVLYANKETSTNTIRPWLKVVNSGSSSIDL 602

QY 492 STVTVRVWFTTRDGSSTLVVNCDAAMCGCNRASFUSVNPATPTAOTYLOLSFTGCT-- 549

Db 603 SRVTIRYWTVDGERAQSAIS-DWAQICASNVTFKFKLSSSVSGADYILEIGFKSCAGO 661

QY 550 LAAGSGSTGEIQNRVYKSDWSNFDENDSY-GTNTAFQDDTKVTIVVYNGRLVWGTPEPSGT 608

Db 662 LQPKDTGEIQIRFNKSDWSNNOGNDWSIQSMTSYGENEKTAYIDGLVMQGEPSGT 721

QY 609 SPSPPTSP-----SPTSPSPSPPTSPSSSPSPSPSPSPSPSPSPSPSPSPSPSPSSSCVG 664

Db 722 TSPSTSTPTVTPT 781

QY 665 CRATYVNSDWSGSGFTATVTV-TNTGSRATSGWTVANSFGCNGQVTVNTWNTALTQSCASV 723

Db 782 SPSPSVVE-----TJNTNAGRTQIS---PYIYGANODIEGVVHSARRLGGCKRL 827

QY 724 TATNLSYN 731

Db 828 TGYNWEEN 835

RESULT 3

T31337

1,4-beta-glucanase (EC 3.2.1.-) - Anaerocellum thermophilum (fragment)

C:Species: Anaerocellum thermophilum

C:Date: 02-Sep-2000 #sequence_revision 02-Sep-2000 #text_change 02-Sep-2000

C:Accession: T31337

R:Zverlov, V.; Mahr, S.; Riedel, K.; Bronnenmeier, K.

Microbiology 144, 457-465, 1998

A:Title: Properties and gene structure of a bifunctional cellulolytic enzyme (CelA)

Omains.

A:Reference number: 221003; MUID:98154434; PMID:9493383

A:Accession: T31337

A>Status: preliminary; translated from CB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-1711 <ZVE>

A:Cross-references: EMBL:286105; NID:e1071329; PID:e350354; PIDN:CAR06786.1

C:Genetics:

A:Gene: celA

C:Keywords: glycosidase; hydrolase

Query Match 16.0%; Score 661.5; DB 2; Length 1711;

Best Local Similarity 25.1%; Pred. No. 1.8e-25;

Matches 222; Conservative 121; Mismatches 245; Indels 295; Gaps 36;

QY 60 GTNNYLYSYQSHAD-----VDDVLAKAQAAMNLSVIRTWGFDIGSLDGSVPITDGNKNGF 114

Db 274 GKDIYKQIIESHLDYWITGYNGERIKYTPKGLAWLDQWGLRYATTATFL-----AF 325

QY 115 YFQYW--DPSTGAPAYNDGPTGLDGLDYATASAAHGLRVIV-----LTN 158

Db 326 VYSDWVGCPTKKEIYR--KFGESQIDYALGSAG-----RSFVVGFGTNPKRPHRTAHS 379

QY 159 DWKEFGGMDQYDK--WYGL-----PYHNFYTDPTQOAYKNWVHLLNRVNSITGVYTKN 212

Db 380 SWADSQISPSYHRTLYGALVGGGSDSYTD-----DISNYVNEVACDYNA 427

QY 213 DPTIFAWELANEPKRCVSGTLP-----TSGCTQATIVNVDMS 252

Db 428 G---FVGLAKAKMOLYGGNPIDPFKAETPTNDEFFVEAGINASGT-----NFIE--- 474

QY 253 AYVKSIDPNHMSVGDGFGYIGTQSGWP-----YNDPSD-----GYDNNALLR 297

Db 475 --TKAIVNNO-----SCWPAKATDKIKFRYFVDLSELKACYSPOGJTL 516

QY 298 VKN-----IDFGTYHLYPN-----YWGO 315

Db 517 STNYNQAKVSGPYVWDASKNIYILVDFTGTLIYPCGQDKYKKEVOFRIAAONVOWDN 576

QY 316 NADWGTONIKDHIAAAAIKPTILEEFGWQTPDRDSVYQVWTQTVRTNGEAGNPNWMLA 375

[illegible]

D	B	226		- - - - - Y N S W S G F Y D E L S W A A V M L Y L A T I N D S T Y L T K A E - - - - - S Y V O N N - - - - - 264
Q	Y	97	I G S L D G S V P T I D G N K - - - - - N C - - - - - F Y F O Y W D P S T 123	
D	B	265	- - - - - P K I S G S N I I D Y K W A H C W D D V H N C A A L L L A K I T D K U T Y K Q I I F S H L I J Y W T T G Y 416	
Q	Y	124	G A P A Y N D G P T G L O G L D - - - - - Y A L A S A A A H G L R V I V V L T N D W K K E F G G M U D Y I K W Y G I J Y 177	
D	B	317	N G E R I K Y T P K G L A W D Q W S L R L A T T T A F - - - L A F V - - - - - Y S D W S G G P - 357	
Q	Y	178	H D N E Y T D P R T O Q A Y K N W N H L L N R V N S I T G V T Y K N D P T I F A M E L A N E P R C V G S G T L I T S G 237	
D	B	358	- - - - - T G K K E T Y R K F G E S O I D Y A L C S T G R S F - - - - - V V G F G T N P P K R 394	
Q	Y	238	T C T Q A T I V N W Y D Q M S A Y V K S I D P N H I W Y S V G D E G F Y I G S T O G S G W P Y N D P S D G V D N N A L I K 297	
D	B	395	P H I R T A H S S W A D S - - - - - Q S T P S Y H R H T L - - Y C A L V G G - P C S D D S Y T D D I I S N V V N E V A C 446	
Q	Y	298	V K N I D F - - - - - G T Y H L Y - - - - - P N Y W C O N A D W G T Q W I K D H I A N A A I G - - - - - K P T I L E R F 343	
D	B	447	D Y N A G E F V G A L A K M Y L L Y G N P I P D F K A I E T P T N D E F F E V A G I N A S C T N F I E I K A I V N N U S 506	
Q	Y	344	C W - - - - - - - - - - - O T P D R D S Y V Q T W O T V R T N G E A G W N - - - - - F W M L A 375	
D	B	507	G W P A R A T N K L K F R Y F V D L S E L I K A G Y S P N O L T I S T N Y N O G A K V S G P Y W D S S R N I Y I L V 566	
Q	Y	376	G N V N G Q P Y P - - - - - - - - - - - N Y D G F N V Y - - - - - Y P S S T A T V L A S E A L I - - - 408	
D	B	567	D F T G T L I Y P G O D K Y K K E V O F R I A A P O N V O M D N S D Y S T O D I K G V S S G S V A K T Y I P L A Y 626	
Q	Y	409	- - - - - - - - - - - S T C T S P - - - - - P S P S S S P S S P S P S P S A S P A S S P S S P S P S 451	
D	B	627	E D I K W M G E E P T G S V S P T P A S V T P T P T P T A T P T P T P T P T P T P T P T P T P T P T P T P T 686	
Q	Y	452	S S P - - - - - - - - - - - V S G V K V Q Y K N D S A P C D N O I K P C L O I V N T G S S V D L S T V T V K Y 498	
D	B	687	S T P T V T P T P T P V S T P A T S G Q I K V L Y A N K E T N S T T N T I R P W L K V N S G S S I D L S R V T I K Y 746	
Q	Y	499	W F T R D G S S T L V Y N C D W A A M C C N I R A S F G S V N A P T P T A D Y I L O L S F T C G T - - L A A G S T 556	
D	B	747	W Y T V D G E R A O S A I S - D W A Q I G A S N V T F K V K L S S V S G A D Y L E I F K R G C A Q O L O P K U T 805	
Q	Y	557	G E I Q R N V K S D W S N F D E T N D Y S - G T N T A P O D M T K V T V V N G R L V M G T P S C T S P T P S 615	
D	B	806	G E I Q R N K D W S N Y N O G N D W S H I O S M T S Y G E N E K V T A I D G V L W G O P S C T T P A P T S T 865	
Q	Y	616	P - - - - - S P T P S P S P T P S S S P S P S P S P S P S P - - - - - S D W S G F T A T V T - - - - - V T N T G S 690	
D	B	866	P T V T V T P T P T P T V T P T V T A T P T P T P T P T P T P T P T P T P T P T P T P T P T P T P T P T P T 925	
Q	Y	653	- - - - - - - - - - - S P S P S V S S G V C C R A T Y V N - - - - - S D W S G F T A T V T - - - - - V T N T G S 690	
D	B	926	I R P W L K V N S G S S I D L S R V T I R Y T V T V D G E R A O S A I S O M A I G A S N V T F K V K L S S S V S 985	
Q	Y	691	R A T S C W T V A W S F G N O T V - - - - - - - - - - - T N Y - - - - - W N T A L T O S G A S V T A T 726	
D	B	986	G A D Y I L E I G F K S G A G O L O P C K D T G E I O I R F N K D D W S N Y N O G N D W S H I O S M T S Y G E N E K V T 1045	
Q	Y	727	N L S P - N N V I O P G O S T T F G F N G S Y S G T N T A P T L C T 760	
D	B	1046	--A Y I D G V L W M G O E P -----S G A T P A P T V T P T 1070	

```

Query Match: 15.3% Score 632.5; DB 2; Length 1742;
Best local similarity 22.8%; Pred. No. 5e-24;
Matches 227; Conservative 143; Mismatches 90; Indels 335; Gaps 37;

Q: 7 PARAFVTAAGTAAVAAATLCGTSITMPSA-----TAAPAGVETASGGGFV 50
      ||||| :||: |
Db 170 PGSAWAETAAASLAASITLVKDRNPTRAAATYLOHAKDLYEAFVETKSDSGYTAANG----- 225

QY 51 LNLGPLYRGYGTNNYYLSYQSHADY-----DDVLAKAAQANNLVSIRTWGFID 96

```

RESULT 5
T31085
xylanase - Caldicellulosiruptor sp.
C:Species: Caldicellulosiruptor sp.
C:Date: 02-Sep-2000 #sequence_revision 02-Sep-2000 #text_change 02-Sep-2000
C:Accession: T31085
R:Morris, D.D.; Gibbs, M.D.; Ford, M.; Thomas, J.; Bergquist, P.L.
Submitted to the EMBL Data Library, December 1997
A:Description: Family 10 and 11 xylanase genes from Caldicellulosiruptor sp.

C>Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Dec-2001
C:Accession: D86153
R:Theologis, A.: Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Creasy, T.H.; Dewar, K.;
ansen, N.E.; Hughes, B.; Huizar, L.
Nature 408, 816-820, 2000
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.;
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziani,
Rizzo, M.; Rooney, T.; Kowley, D.; Sakano, H.
ker, M.; Salzborg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A:Reference number: A86141; MUID:21016719; PMID:11130712
A:Accession: D86153
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-411 <SPO>
A:Cross-references: GB:AE005172; NID:99857528; PIDN:AAC00883.1; GSPDB:GN00141
C:Genetics:
A:Map position: 1

Query Match 12.8%; Score 526.5; DB 2; Length 411;
Best Local Similarity 32.0%; Pred. No. 1.9e-19;
Matches 133; Conservative 70; Mismatches 142; Indels 71; Gaps 14;
Qy 36 AAPAGFVTASGGQFVLNGLPYRGCTNNYL-----SYQSHADVDDVLAKAQAAMNLSV 88
Db 24 AVKTGFVGRNCTQFVLNCEQVYLNGFNAYNMMTTAAATASKGRATVTTALRQASAVGMNV 83
Qy 89 IRTWGFIDIGSLDGSVPTIDGNKNGFFQYWDPTGAP-AYNDGPTGLQGLDYAIAASAAA 147
Db 84 ARINGF-----NEGDIY-----PLQISPGSYSEDV--FKGLDFVYVEAGR 121
Qy 148 HGLRVIVLTLNDKKEFGMDQYDKWYGLPYHDNFYTDPTQOAYKNVNHLLNRVNSITG 207
Db 122 FNKLIIISLVNNEEDYGRKKYKWEAGLDEPDEYTHNSAVKQFYKNHVTVLTAKNTITG 181
Qy 208 VTKNDPTIFAWELANEPKRCVSGTLPSTGCTQATIV-NVVDQMSAYVKSIDPNHNVSV 266
Db 182 RMYKDDPTIFSWELINEPRCND-----TASNILQDVKWEMASVYKSIDSNHLEI 232
Qy 267 GDEGFYIGSTOGSOWPN-----DPSDGVNALLRVKNIIDFTYHLYPNYW----- 313
Db 233 GLEGFY-GESIPERTVYNGRVLGTGDFITNN--QIPDIDFATIHYPDSWLPLOSSR 288
Qy 314 -GNQADMGCTQWIKDHANA-AAIGKPTILEEF-----GWQTPDRDSVYQTWTQVTRN 364
Db 289 TGEODTFVDRWGAHIEDCDNIKKPLITEFGKSSKYPGFSLEKRNKFFQRYVDVIYDS 348
Qy 365 GEAGWN-----FWMAGNVNGQPNYVDGNVYVPSSTATVLAASEALAIETGTSP 415
Db 349 ARAGGCTCGVFWOLTNRCTCLLDGDEVEFMQAGPNITTAQLIADOSSKLNKLYPP 404

RESULT 8
T48214
end-1,4-beta-mannosidase-like protein - Arabidopsis thaliana
N:Alternate names: protein T20L15.200
C:Species: Arabidopsis thaliana (mouse-ear cress)
C>Date: 20-Apr-2000 #sequence_revision 20-Apr-2000 #text_change 20-Apr-2000
C:Accession: T48214
R:Bevan, M.; Peters, S.A.; van Staveren, M.; Dirkse, W.; Stiekema, W.; Bancroft, I.; Mey
submitted to the Protein Sequence Database, March 2000
A:Reference number: 224488
A:Accession: T48214
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-448 <BEV>
A:Cross-references: EMBL:AL162351
A:Experimental source: cultivar Columbia; BAC clone T20L15
C:Genetics:
A:Map position: 5
A: Intron: 34/2; 126/3; 191/3; 303/2

A:Note: T20L15.200

Query Match 12.1%; Score 498; DB 2; Length 448;
Best Local Similarity 32.2%; Pred. No. 5.4e-18;
Matches 132; Conservative 66; Mismatches 130; Indels 82; Gaps 17;
Qy 42 VTASGGQFVLNGLPYRGCTNNYL-----SYQSHADVDDVLAKAQAAMNLSVIRTWGTFID 96
Db 48 VORKGMOFTLNGQPFYVNGFNMTTAAADNSTRGKVTVEVPOQASAVGTMVGRWAFND 107
Qy 97 IGSLOGSVPTIDGNKNGFFQYWDPTGAPAYNDGPTGLQGLDYAIAASAAHGLRVIVL 156
Db 108 -----QWRALQKSPSYDEEV-FKALDFVLSLEARKYKIRLILSL 146
Qy 157 TNDKKEFGMDQYDKW-----YGLPYHDNFYTDPTQOAYKNVNHLLNRVNSITGVITYK 211
Db 147 VNNDAVGGKAYVKGWNASGILNTSDDDFTNTRNFYQSHVTVLUNKVNTFNITTYK 206
Qy 212 NDPTIFAWELANEPKRCVSGTLPSTGCTQATIVNVVDQMSAYVKSIDPNHNVSVGDEGF 271
Db 207 NDPTIFAWELANEPKRCPSD-----PSGDKLQ-----SWIQEMAVFVKSLDKHLVEICLEGF 258
Qy 272 YIGS-----TOGSGWPN-----NDPSDGVNALLRVKNIIDFTYHLYPNYWGONA-----IW 319
Db 259 YGSPAPARTRENPNPYAAQVGTDFIRNQVL---GIDFASVIVYPDSWISPAVSNSFLEF 315
Qy 320 GTOWIKDHANAAA-IGKPTILEEF-----GWQTPDRD-----SVYOTWTQTVRTNGEA 367
Db 316 TSSWQAHVEADAEMVGLMPVLTFFGVSAHDPGNTSFRDMLNTVYKMTLNSTRKGG-A 374
Qy 368 GWNFWMLAGNVNGQPN-----YDGFNVY-----YPSSTATVLAASEALAI 408
Db 375 G-----AGSLVQVFOGAEFMDGQYAVYLTRAHTASKIISLOSKRLAI 418

RESULT 9

T09048

probable mannan endo-1,4-beta-mannosidase (EC 3.2.1.78) - Arabidopsis thaliana
N:Alternate names: (1-4)-beta-mannan endohydrolase; protein F26K10.200
C:Species: Arabidopsis thaliana (mouse-ear cress)
C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
R:Bevan, M.; Lennard, N.; Quail, M.; Harris, B.; Rajandream, M.A.; Barrell, B.G.; Ban
submitted to the Protein Sequence Database, June 1999
A:Reference number: 216533

A:Accession: T09048

A:Molecule type: DNA

A:Residues: 1-431 <BEV>

A:Cross-references: EMBL:AL049803; ATSP:F26K10.200; GSPDB:GN00062

A:Experimental source: cultivar Columbia; BAC clone F26K10

C:Genetics:

A:Gene: ATSP:F26K10.200

A:Map position: 4

A: Intron: 120/3; 186/3; 227/3; 298/2

C:Keywords: glycosidase; hydrolase

Query Match 12.0%; Score 495; DB 2; Length 431;
Best Local Similarity 32.9%; Pred. No. 7.3e-18;
Matches 124; Conservative 55; Mismatches 124; Indels 74; Gaps 16;

Qy 37 APAGFVTASGGQFVLNGLPYRGCTNNYLSYQSHAD-----VDDVLAKAQAAMNLSV 89

Db 37 AKLGFVVRNGTQFVDDKPLVYVGNWSYW--FMDHAYDEHSRNLVGEHLEACAKMGLTVC 94

Qy 90 RTWGFIDIGSLDGSVPTIDGNKNGFFQYWDPTGAPAYNDGPTGLQGLDYAIAASAAH 149

Db 95 RTWAF-----NDGCYNALQI-----SPGRFDERV-FOALDHDVIAEARKHD 133

Qy 150 LRVIVLTLNDKKEFGMDQYDKW-----YGL-PYHDNFYTDPTQOAYKNVNHLLNRVN 203

Db 134 VRLLSLVNLNQAYGGKTYVYKWAQEGVGLSSNDSEFFDPDSIRNFYKNLYKLLTRKN 193

Qy 204 SITGVTYKNDPTIFAWELANEPKRCVSGTLPSTGCTQATIVNVVDQMSAYVKSIDPNH 263

Db 194 SVTGLEYNODPTIFAWELINEPRC-----TTDVSQ-----KTLDWDIDENTGFIKSIDOKHL 245
 Qy 264 VSVGDEGFIYTGSTGSC-----WPNNDPSGDVNDNALLKVKNIIDFGTYHLYPNYWGON 316
 Db 246 LTVGLEGFY-CPNSPKGLTVNPEQWASOLGTFDFQNS---NSSNIDFASVHIYPDIHFHN 301
 Qy 317 ADWG-----TQWTKDHTANA-AAIGKPTILEERG-----WQTPDRDSVYO-----TW 357
 Db 302 QTFFEEKLVVVKWQSHLEDGLKELKPKVLPTEFGLSNQNKDYEPSQDKFYRIIFDVVY 361
 Qy 358 TQTVRTNGEAGNFMML 374
 Db 362 KSAKKRSGACTLVWLQ 378

RESULT 10

S36859

c1pA protein - Clostridium thermocellum

N:Alternate names: probable cellulosome protein large chain SL

C:Species: Clostridium thermocellum

C:Date: 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 01-Dec-2000

C:Accession: S36859; S33527; S25767; S28659; T18261

R:Gerngross, U.T.; Demain, A.L.

A:Reference number: S36859

A:Accession: S36859

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-1854 <GER>

R:Gerngross, U.T.; Romaniec, M.P.M.; Kobayashi, T.; Huskisson, N.S.; Demain, A.L.

Mol. Microbiol. 8, 325-334, 1993

A:Title: Sequencing of a Clostridium thermocellum gene (c1pA) encoding the cellulosomal

A:Reference number: S33527; MUID:93302508; PMID:8316083

A:Accession: S33527

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-1615,1617-1854 <GE2>

A:Cross-references: EMBL:L08665

R:Fujino, T.; Beguin, P.; Aubert, J.P.

FEMS Microbiol. Lett. 94, 165-170, 1992

A:Title: Cloning of a Clostridium thermocellum DNA fragment encoding polypeptides that b

A:Reference number: S25767

A:Accession: S25767

A:Molecule type: DNA

A:Residues: 1216-1232,'Y',1234-1241,'T',1243-1319,'R',1321-1615,1617-1854 <FUJ>

A:Cross-references: EMBL:X67406

R:Bequin, P.

submitted to the EMBL Data Library, August 1992

A:Reference number: S28659

A:Accession: S28659

A:Molecule type: DNA

A:Residues: 1216-1232,'Y',1234-1241,'T',1243-1319,'R',1321-1615,1617-1768,'R',1770-1854

A:Cross-references: EMBL:X67406

R:Fujino, T.; Beguin, P.; Aubert, J.P.

J. Bacteriol. 175, 1891-1899, 1993

A:Title: Organization of a Clostridium thermocellum gene cluster encoding the cellulosome

e.

A:Reference number: Z18847; MUID:93209931; PMID:8458832

A:Accession: T18261

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1821-1854 <FU2>

A:Cross-references: EMBL:X67506; NID:g296879; PID:g296880; PIDN:CAA47840.1

C:Genetics:

A:Gene: c1pA

Query Match

Best Local Similarity 25.1%; Pred. No. 1.4e-16;

Matches 165; Conservative 102; Mismatches 208; Indels 182; Gaps 24;

Qy 229 GSGTLPTSGTGTQATIVNWDMSA-----YKSIDPNHMSVSGDE----GFYIGSTQG 278

Db 114 GRGTYAITODGVFATIVATVKSAAAAAPITLLEVCAFADNDLVEISTTFVACGVNLCSSVP 173
 Qy 279 SCWPNYNDPSGDVNDNALLRVKNI--DFGTYHLYPNYWGONADMGQWQIWKHIANAAAI CK 336
 Db 174 TTQP--NVPSCGV---VVEICKVTGVSCTTVEIPVYFRGVPSKG-----IANDDFVFR 221
 Qy 337 --PTILEEFG-----WQT---PDR-----DSVYQT 356
 Db 222 YDPNVYLEITIGIDPGDIIVDPNPTKSFDTAIYPRDKIIVLFAEDSGTGCAYAITKDVFAK 281
 Qy 357 WTQTVRTNGEAGNFMWLAGNVNGOPYNYDGFNVYVPSSTATVLAASEALISTCT---S 413
 Db 282 IRATVKSSA-----PGYITFDEVGGFAONDLYEOKVSVFIDKGVNWN 323
 Qy 414 PPPSPSSSPSSP 473
 Db 324 ATPTKGATPTNTATPTKSAATATPTRPSVPTNTPTNTANTPVSGNLKVEFYNSNP-SDITN 384
 Qy 474 QIKPCLQLVNTGSSSVLDLSTVTVRYWFTTRDGSSTLVVYNCDAAM-----GCGNJK 524
 Db 384 SINPQFVNTGSSAIDLSKLTLYRYITVDGOKDTFW-CDHAAIIGNSGYNIGTISNVK 442
 Qy 525 ASFGSVNPATPTADTYLQLSFTGTLAAGSTGTEIQNRVKNKSDMSNFDDETNDYNYGTNTA 584
 Db 443 GTFVKMSSTNNADTYLEISFTGCTLEPCAIV-QIOGRFAKNDWSNYTOSNDYSFKSASQ 501
 Qy 585 FQDWTKVTVVYVNGRLVWCTEPSCTSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSP 644
 Db 502 FVEWDQVTAYLVNGLVMGKEPG---SVVPSTQPVTTT-PATTKPPATTKPATTPISD 557
 Qy 645 TP-----SPSPSPSPSPSPSSSCV-----GCRATY-----VV 671
 Db 558 DPNAIKTKVDTVNAKPGDTWIPVRFSGIPSKGIANCDFVSYSDPNVLEIEIKPCELIIV 617
 Qy 672 NSDMGSGF-----TATVTVTNTGSRATSGTWAVWSEFGNQTVTNY 711
 Db 618 DPNDKSFDTAVYPDRKIIVFLFAEDSCTGAYAITKDCVFATI----- 660
 Qy 712 WNTALTOSGAS-----VTATNLSYNNVLOPGQSTTFGFGNSGYNIGTNTAPTICT 760
 Db 661 --VAKVKGAPNGLSVIKFVEVGFGFANNDLVE--QRTOFFDGVNNGDVTPTPTTT 713

RESULT 11

G69593

cellulase (EC 3.2.1.4) bgIC precursor - Bacillus subtilis

N:Alternate names: endo-1,4-beta-glucanase

C:Species: Bacillus subtilis

C:Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 21-Jul-2000

C:Accession: G69593; A26114; I40353; S24239; S49103; I39803

R:Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Ber

C.: Bron, S.; Brouillet, S.; Brusch, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.;

A.: Ehrlich, S.D.; Emerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari,

Nature 390, 249-256, 1997

A:Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Gal

lech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo,

Koetter, P.; Koningsstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinc

A:Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Mau

Y, M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portete

Rieger, M.; Rivolta, C.; Rocha, E.; Roche, M.; Rose, M.; Sadale, Y.; Sato, T.; Scanl

A:Authors: Schleich, S.; Schroeder, R.; Scofield, F.; Sekiguchi, J.; Sekowska, A.; Se

akeuchi, M.; Tamakoshi, A.; Tanaka, T.; Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiya

T.: Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida

A:Authors: Yoshikawa, H.F.; Zumstein, E.; Yoshikawa, H.; Danchin, A.

A:Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtili

A:Reference number: A69580; MUID:98044033; PMID:9384377

A:Accession: G69593

A:Status: nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-508 <KUN>

A:Cross-references: GB:Z99113; GB:AL009126; NID:g2634090; PIDN:CAB13696.1; PID:e18134

A:Experimental source: strain 168

R: Mackay, R.M.; Lo, A.; Willick, G.; Zuker, M.; Baird, S.; Dove, M.; Moranelli, F.; Sell
Nucleic Acids Res. 14, 9159-9170, 1986
A: Title: Structure of a *Bacillus subtilis* endo-beta-1,4-glucanase gene.
A: Reference number: A26114; MUID: 87066783; PMID: 3024130
A: Accession: A26114
A: Molecule type: DNA
A: Residues: 10-508 <MAC>
A: Cross-references: GB:X04689; NID: g39823; PIDN: CAA28392.1; PID: g39824
A: Experimental source: strain PAP115
A: Note: part of this sequence, including the amino end of the mature form, was confirmed
R: Lindahl, V.; Aa, K.; Tronsmo, A.
Antonie Van Leeuwenhoek 66, 327-332, 1994
A: Title: Nucleotide sequence of an endo-beta-1,4-glucanase gene from *Bacillus subtilis* O
A: Reference number: 140353; MUID: 95225656; PMID: 7710280
A: Accession: 140353
A: Accession: translated from GB/EMBL/DDBJ
A: Status: preliminary
A: Molecule type: DNA
A: Residues: 10-291, 'N', 293-508 <LIN1>
A: Cross-references: EMBL: X67044; NID: g39776; PIDN: CAA47429.1; PID: g39777
R: Lindahl, V.; Aa, K.
submitted to the EMBL Data Library, June 1992
A: Reference number: S24239
A: Accession: S24239
A: Molecule type: DNA
A: Residues: 10-291, 'N', 293-508 <LIN2>
A: Cross-references: EMBL: X67044; NID: g39776; PIDN: CAA47429.1; PID: g39777
A: Experimental source: strain CK-2
R: Wolf, M.; Geczi, A.; Borriass, R.
submitted to the EMBL Data Library, December 1993
A: Description: Genes encoding beta glucan-hydrolyzing enzymes in *Bacillus subtilis*: cons
A: Reference number: S49103
A: Accession: S49103
A: Molecule type: DNA
A: Residues: 10-508 <MOL2>
A: Cross-references: EMBL: 229076; NID: g509266; PIDN: CAA82317.1; PID: g509267
R: Seo, Y.S.; Lee, Y.H.; Pak, U.H.; Kang, H.
Korean J. Microbiol. 24, 236-242, 1986
A: Title: Analysis on the nucleotide sequence of the signal region of *Bacillus subtilis*
A: Reference number: I39803
A: Accession: I39803
A: Status: preliminary; translated from GB/EMBL/DDBJ
A: Molecule type: DNA
A: Residues: 10-13, 'V', 15, 'N', 17-18, 'V', 20-21, 'F', 23, 'A', 25-26, 'AI', 29-31, 'P', 33, 'PQ', 36, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 100, 101, 102, 103, 104, 105, 106, 107, 108, 109, 110, 111, 112, 113, 114, 115, 116, 117, 118, 119, 120, 121, 122, 123, 124, 125, 126, 127, 128, 129, 130, 131, 132, 133, 134, 135, 136, 137, 138, 139, 140, 141, 142, 143, 144, 145, 146, 147, 148, 149, 150, 151, 152, 153, 154, 155, 156, 157, 158, 159, 160, 161, 162, 163, 164, 165, 166, 167, 168, 169, 170, 171, 172, 173, 174, 175, 176, 177, 178, 179, 180, 181, 182, 183, 184, 185, 186, 187, 188, 189, 190, 191, 192, 193, 194, 195, 196, 197, 198, 199, 200, 201, 202, 203, 204, 205, 206, 207, 208, 209, 210, 211, 212, 213, 214, 215, 216, 217, 218, 219, 220, 221, 222, 223, 224, 225, 226, 227, 228, 229, 230, 231, 232, 233, 234, 235, 236, 237, 238, 239, 240, 241, 242, 243, 244, 245, 246, 247, 248, 249, 250, 251, 252, 253, 254, 255, 256, 257, 258, 259, 260, 261, 262, 263, 264, 265, 266, 267, 268, 269, 270, 271, 272, 273, 274, 275, 276, 277, 278, 279, 280, 281, 282, 283, 284, 285, 286, 287, 288, 289, 290, 291, 292, 293, 294, 295, 296, 297, 298, 299, 300, 301, 302, 303, 304, 305, 306, 307, 308, 309, 310, 311, 312, 313, 314, 315, 316, 317, 318, 319, 320, 321, 322, 323, 324, 325, 326, 327, 328, 329, 330, 331, 332, 333, 334, 335, 336, 337, 338, 339, 340, 341, 342, 343, 344, 345, 346, 347, 348, 349, 350, 351, 352, 353, 354, 355, 356, 357, 358, 359, 360, 361, 362, 363, 364, 365, 366, 367, 368, 369, 370, 371, 372, 373, 374, 375, 376, 377, 378, 379, 380, 381, 382, 383, 384, 385, 386, 387, 388, 389, 390, 391, 392, 393, 394, 395, 396, 397, 398, 399, 400, 401, 402, 403, 404, 405, 406, 407, 408, 409, 410, 411, 412, 413, 414, 415, 416, 417, 418, 419, 420, 421, 422, 423, 424, 425, 426, 427, 428, 429, 430, 431, 432, 433, 434, 435, 436, 437, 438, 439, 440, 441, 442, 443, 444, 445, 446, 447, 448, 449, 450, 451, 452, 453, 454, 455, 456, 457, 458, 459, 460, 461, 462, 463, 464, 465, 466, 467, 468, 469, 470, 471, 472, 473, 474, 475, 476, 477, 478, 479, 480, 481, 482, 483, 484, 485, 486, 487, 488, 489, 490, 491, 492, 493, 494, 495, 496, 497, 498, 499, 500, 501, 502, 503, 504, 505, 506, 507, 508, 509, 510, 511, 512, 513, 514, 515, 516, 517, 518, 519, 520, 521, 522, 523, 524, 525, 526, 527, 528, 529, 530, 531, 532, 533, 534, 535, 536, 537, 538, 539, 540, 541, 542, 543, 544, 545, 546, 547, 548, 549, 550, 551, 552, 553, 554, 555, 556, 557, 558, 559, 560, 561, 562, 563, 564, 565, 566, 567, 568, 569, 570, 571, 572, 573, 574, 575, 576, 577, 578, 579, 580, 581, 582, 583, 584, 585, 586, 587, 588, 589, 590, 591, 592, 593, 594, 595, 596, 597, 598, 599, 600, 601, 602, 603, 604, 605, 606, 607, 608, 609, 610, 611, 612, 613, 614, 615, 616, 617, 618, 619, 620, 621, 622, 623, 624, 625, 626, 627, 628, 629, 630, 631, 632, 633, 634, 635, 636, 637, 638, 639, 640, 641, 642, 643, 644, 645, 646, 647, 648, 649, 650, 651, 652, 653, 654, 655, 656, 65

Db 116 SVKNKVEAAEAKELGIYVII-----DW-----HILNDGNPNQ-NK 151
QY 193 NWVHLLNRVNSITGVYTKNDPTIFAMELANEPRCVSGTLPTSGTCTQATIVNWVDS 252
Db 152 EKAKFPEKSSSL-----YGNTPNVI-YEIANEP-----NGD-----VNKKRDIK 190
QY 253 AYVKSI-----DPNHVSVGDEGFYIGSTQSGWPNYDPSGDVNNALLRVKNIDFGT 305
Db 191 PYAEVVISVIRKNDPDNIIV-----CTGTWSQDVNDAAADQ--LKDANVMY-A 236
QY 306 YHLYPNWGNADWGTOWIKDHIANAAAIKPTILEEFGHQTDPDRDSVYQVOTVTRNG 365
Db 237 LIHFYAGTHGOS-----LRDKANYALSKGAPIFVTE--WGTS--ASG 274
QY 366 EAG-----WNFMLAGVNVGQPNYDGFNVYPSSTATVLAASEALAICTCTSPPPS 417
Db 275 NGGVFLDOSREMLNYLDSKNIS---WVNNLSDKQESSALKPCASK-----TGCWPLTD 326
QY 418 PSSSPSSPSSP 477
Db 327 LTASGTFVRENIRKTKUSTKDVPEPAPQDNPTQEK-----GVSVOYKAGDGRVNSNOIRP 381
QY 478 GLOLVNTGSSVDLSTVTYRYWF--TRDGGSTLVYNCDDWAAACGNIIRASFGSVNPATPT 536
Db 382 QLIHKNNGNATVDLKDVTARTYNNKNGON---FDCDYAOMCGNLTTHKFVTLHKPKQG 438
QY 537 ADTYLQLSFTGGTLAAGSTGEIQNRVKNKSDWNSFDETDNDYSYGTNTAFQDWTKVTVVN 596
Db 439 ADTYLELGFKTGTLPCASTGNIOLRHNDWNSYAGSDYSFFQSNFTKTKITLYHQ 498
QY 597 GRLVWGTEPS 606
Db 499 GKLWGTPEP 508

RESULT 13
A27198
cellulase (EC 3.2.1.4) precursor - Bacillus subtilis (strain IF03034)
N:Alternate names: endo-1,4-beta-glucanase
C:Species: Bacillus subtilis
C:Date: 15-Dec-1988 #sequence_revision 15-Dec-1988 #text_change 15-Oct-1999
C:Accession: A27198
R:Nakamura, A.; Uozumi, T.; Beppu, T.
Eur. J. Biochem. 164, 317-320, 1987
A:Title: Nucleotide sequence of a cellulase gene of Bacillus subtilis.
A:Reference number: A27198; MUID:87190397; PMID:3106035
A:Accession: A27198
A:Molecule type: DNA
A:Residues: 1-499 <NAK>
A:Cross-references: GB:M28332; NID:gl42670; PIDN:AAA22307.1; PID:gl42671
A:Experimental source: strain IF03034
C:Function:
A:Description: hydrolysis of 1,4-beta-D-glucosidic linkages in beta-D-glucans such as ce
A:Pathway: cellulose degradation
C:Keywords: extracellular protein; glycosidase; hydrolase; polysaccharide degradation
F:1-36/Domain: signal sequence #status predicted <SIG>

Query Match 11.0%; Score 451.5; DB 2: Length 499;
Best Local Similarity 25.3%; Pred. No. 1.3e-15;
Matches 155; Conservative 99; Mismatches 211; Indels 147; Gaps 27;

QY 12 VATAAGTAVAAATLCSITHPSATAAPAGVFTASGGQFV-LNGLPYRYGGTNNYLSYQS 70
Db 16 VLTMGGLQASPASAACTKT-----PAKNGOLSITKQTLVNRDQKAVOLKGISSHGLQWYG 71
QY 71 HADVDVLAQAAMNLSVIR-TWGFIDIGSLDGSVPTIDCKNKGFFQYWDPTSGAPAYN 129
Db 72 -----DFVNKD---SLKWLDDNG---ITVFRAMTYTADG-----GYI 103
QY 130 DPTGLQGLDYAIAASAAHGLRIVVLTNDWKEFGMDQYDKWGLPYHDFNTDPRTOQ 189
Db 104 DNPVSKNKVKEAAEAKELGIYVII-----DW-----HILNDGNPNQ 140

QY 190 AYKNWVHLLNRVNSITGVYTKNDPTIFAMELANEPRCVSGTLPTSGTCTQATIVNWVD 249
Db 141 -HREKADKFFEKSSSL-----YGNTPNVI-YEIANEP-----NGD-----VNKKR 178
QY 250 OMSAYVKS-----DPNHVSVGDEGFYIGSTQSGWPNYDPSGDVNNALLRVKNID 302
Db 179 DIKPYAEVVISVIRKNDPDNIIV-----CTGTWSQDVNDAAADQ--LKDANVM 225
QY 303 FGYTHLYPNWGNADWGTOWIKDHIANAAAIKPTILEEFGHQTDPDRDSVYQVOTVTR 362
Db 226 Y-ALHFYAGTHGOS-----LRDKANYALSKGAPIFVTE--WGTS-- 262
QY 363 TNGEAG-----WNFMLAGVNVGQPNYDGFNVYPSSTATVLAASEALAICTGTSP 414
Db 263 ASNGGVFLDOSREMLNYLDSKNIS---WVNNLSDKQESSALKPCASK-----TCGWP 314
QY 415 PPSPPSSPSSP 474
Db 315 LTDLTASGTFVRENILGNKDKSTKERPETPAQDNPAQEN-----GISVOYKAGDGVNSN 369
QY 475 IKPGLQLVNTGSSVDLSTVTYRYWF--TRDGGSTLVYNCDDWAAACGNIIRASFGSVNPA 533
Db 370 IRPOLHTKNNGNATVDLKDVTARTYNNKNGON---FDCDYAOMCGNLTTHKFVTLHKP 426
QY 534 TPTADTYLQLSFTGGTLAAGSTGEIQNRVKNKSDWNSFDETDNDYSYGTNTAFQDWTKV 593
Db 427 KQAGADTYLELGFKTGTLPCASTGNIOLRHNDWNSYAGSDYSFFQSNFTKTKITL 486
QY 594 YVNGRLVWGTEP 605
Db 487 YHOGKLIWGTEP 498

RESULT 14
S54744
cellulase (EC 3.2.1.4) CelV1 precursor - Erwinia carotovora (SCC 3193)
N:Alternate names: endo-1,4-beta-glucanase
C:Species: Erwinia carotovora
A:Variety: SCC 3193
C:Date: 27-Oct-1995 #sequence_revision 03-Nov-1995 #text_change 08-Oct-1999
C:Accession: S54744; S4996
R:Maee, A.; Heikkinen, R.; Palva, E.T.
Mol. Gen. Genet. 247, 17-26, 1995
A:Title: Structure and regulation of the Erwinia carotovora subspecies carotovora SCC
A:Reference number: S54744; MUID:95231512; PMID:7715600
A:Accession: S54744
A:Molecule type: DNA
A:Residues: 1-504 <MAE>
A:Cross-references: EMBL:X79241; NID:g493492; PIDN:CAA55823.1; PID:g493493
C:Genetics:
A:Gene: celV1
C:Function:
A:Description: hydrolysis of 1,4-beta-D-glucosidic linkages in beta-D-glucans such as
A:Pathway: cellulose degradation
C:Keywords: glycosidase; hydrolase; polysaccharide degradation
F:1-32/Domain: signal sequence #status predicted <SIG>
F:33-504/Product: cellulase #status predicted <MAT>

Query Match 10.9%; Score 451; DB 2: Length 504;
Best Local Similarity 29.0%; Pred. No. 1.4e-15;
Matches 142; Conservative 64; Mismatches 158; Indels 126; Gaps 24;

QY 143 ASAAHGLRIVVLTNDWKEFGMDQYDKWGLPYHDFNTDPRTOOAYKNWVHLLNRV 202
Db 114 AVAAAGLGVYII--DWHTLS-----DN-----YKAAKIFFAEM 151
QY 203 NSITGVYTKNDPTIFAMELANEPRCVSGTLPTSGTCTQATIVNWVDSMAY-----V 255
Db 152 AGL-----YGNSPNVI-YEIANEP--GS-----VTWNCQIRPYALEVTDTI 190
QY 256 KSIDPNHVSVDGEGFYIGSTQSGWPNYDPSDGVN-----NALLRVKNIDFGTYHLYP 310
Db 191 RSKDPONLIIV-----GSGTWSQDIHDAADNQLPDPNTLYAL-----HFA 231

Search completed: November 13, 2002, 11:57:31
Job time : 41.5819 secs

GenCore version 5.1.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 13, 2002, 11:47:41 : Search time 17.5172 seconds
(without alignments)
1804.220 Million cell updates/sec

Title: US-09-917-378-1

Perfect score: 4122

Sequence: 1 MGLVRRPARAFVATAAGTAV.....GFNGSYSGTNTAPTTLCTAS 762

Scoring table: BIOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_40.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	696.5	16.9	562	1 GUN1_ACICE	P54583 acidothermu
2	676.5	16.4	1331	1 MANB_CALSA	P22533 caldocellum
3	632.5	15.3	1742	1 GUNB_CALSA	P22534 caldocellum
4	603.5	14.6	1039	1 GUNB_CALSA	P10474 c endogluca
5	484	11.7	1853	1 C1PA_CLOTH	Q06851 clostridium
6	476	11.5	772	1 C1PB_CLOTH	Q01866 clostridium
7	460	11.2	499	1 GUN2_BACSU	P10475 bacillus su
8	454	11.0	499	1 GUN1_BACSU	P07983 bacillus su
9	451	10.9	504	1 GUNW_ERWCA	Q59395 erwiniia car
10	451	10.9	747	1 GUNW_CELFI	P50400 cellulomona
11	447	10.8	499	1 GUN3_BACSU	P23549 bacillus su
12	445	10.8	700	1 GUNA_PAELE	P29719 paenibacilli
13	426.5	10.3	505	1 GUNV_ERWCA	Q47096 erwiniia car
14	394	9.6	145	1 YCEA_PAELE	P29718 paenibacilli
15	380.5	9.2	456	1 GUNA_MICBI	P26414 microbispor
16	380	9.2	441	1 GUN2_THEFU	P26222 thermomonos
17	376.5	9.1	880	1 GUN4_THEFU	P10476 pseudomonas
18	362	8.8	962	1 GUNA_PSEFL	P26225 cellulomona
19	335	8.1	1045	1 GUNB_CELFI	P50401 cellulomona
20	330	8.0	872	1 GUXA_CELFI	P50900 clostridium
21	315	7.6	914	1 GUX2_CLOS	P50899 cellulomona
22	313.5	7.6	1090	1 GUXB_CELFI	P29786 cellulomona
23	311.5	7.6	484	1 GUX_CELFI	P23659 clostridium
24	308.5	7.5	986	1 GUNZ_CLOS	O10341 oryza pseu
25	300	7.3	279	1 Y091_NPVOP	Q02934 xanthomonas
26	298	7.2	879	1 GUN1_CLOTH	P19487 xanthomonas
27	287	7.0	484	1 GUNA_XANCP	P19487 xanthomonas
28	285.5	6.9	474	1 VTP3_TTVIV	P19487 xanthomonas
29	274	6.6	555	1 GPI_CHURE	Q9fpg6 chlamydomon
30	268	6.5	515	1 GUNZ_CLOCL	P28623 clostridium
31	252.5	6.1	1229	1 N121_HUMAN	Q9y2n3 homo sapien
32	245	5.9	1367	1 AMYH_YEAST	P08640 saccharomyc
33	245	5.9	1848	1 CBPA_CLOCL	P38058 clostridium

ALIGNMENTS

RESULT 1

GUN1_ACICE	STANDARD;	PRT;	562 AA.
AC	P54583;		
DT	01-OCT-1996 (Rel. 34, Created)		
DT	01-OCT-1996 (Rel. 34, Last sequence update)		
DT	01-OCT-1996 (Rel. 34, Last annotation update)		
DE	Endoglucanase E1 precursor (EC 3.2.1.4) (Endo-1,4-beta-glucanase E1)		
DE	(Cellulase E1) (Endocellulase E1).		
OS	Acidothermus cellulolyticus.		
OC	Bacteria; Actinobacteria; Actinobacteriia (class); Actinobacteridae;		
OC	Actinomycetales; Frankineae; Acidothermaceae; Acidothermus.		
OX	NCBI_TaxID=28049;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN-ATCC 43068 / 11b;		
RA	Laymon R.A., Himmel M.E., Thomas S.R.;		
RL	Submitted (AUG-1995) to the EMBL/GenBank/DBJ databases.		
RN	[2]		
RP	X-RAY CRYSTALLOGRAPHY (2.4 ANGSTROMS) OF 42-398.		
RX	MEDLINE=96346058; PubMed=8718854;		
RA	Sakon J., Adney W.S., Himmel M.E., Thomas S.R., Kaplus P.A.;		
RT	"Crystal structure of thermostable family 5 endocellulase E1 from Acidothermus cellulolyticus in complex with cellobiose."		
RL	Biochemistry 35:10648-10660(1996).		
CC	-!- FUNCTION: THERMOSTABLE ENZYME WITH AN OPTIMAL TEMPERATURE OF 81 DEGREES CELSIUS. HAS A VERY HIGH SPECIFIC ACTIVITY ON CARBOXYMETHYLCELLULOSE.		
CC	-!- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic linkages in cellulose.		
CC	-!- SIMILARITY: BELONGS TO CELLULASE FAMILY A (FAMILY 5 OF GLYCOSYL HYDROLASES).		
CC	-----		
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).		
CC	-----		
CC	EMBL: U33212; AAA75477.1; -		
DR	PDB; LECE; 14-OCT-96.		
DR	InterPro: IPR001919; Bac_cellose-bind.		
DR	InterPro: IPR001547; GH_5.		
DR	Pfam: PF00150; cellulase; 1.		
DR	Pfam: PF00553; CBM_2; 1.		
DR	PROSITE: P500659; GLYCOSYL_HYDROL_F5; 1.		
KW	Cellulose degradation; Hydrolase; Glycosidase; Signal; 3D-structure.		
FT	SIGNAL 1 41		
FT	CHAIN 42 562		ENDOGLUCANASE E1.
FT	DOMAIN 42 400		CATALYTIC.
FT	DOMAIN 401 461		PRO/SER/THR-RICH (LINKER).
FT	DOMAIN 462 562		CELLULOSE-BINDING (BY SIMILARITY).
FT	ACT_SITE 203 203		PROTON DONOR.
FT	ACT_SITE 323 323		NUCLEOPHILE.

34	244.5	5.9	444	1	GUNW_ERWCA	Q59394 erwiniia car
35	243.5	5.9	532	1	SPG2_DICD1	P22698 dictyospori
36	231	5.6	748	1	GUNC_PSEFL	P27033 pseudomonas
37	223.5	5.4	1852	1	RPB1_CAEEL	P16356 raeniorhabdi
38	218.5	5.3	5179	1	MUC2_HUMAN	Q02817 homo sapien
39	218	5.3	619	1	CHIT_STRL1	P36909 streptomyce
40	217	5.3	644	1	XYND_CELFI	P54865 cellulomona
41	216	5.2	534	1	APG-ARATH	P40602 arabidopsis
42	213.5	5.2	449	1	GUNA_CELFI	P07984 cellulomona
43	210.5	5.1	1664	1	SLP1_CLOTH	Q06852 clostridium
44	210	5.1	610	1	CHIT_STRPL	P11220 streptomyce
45	207.5	5.0	535	1	SPKC_SYNY3	P74745 synochocyst


```
Query Match 16.4%; Score 676.5; DB 1; Length 1331;
Best Local Similarity 24.5%; Pred. No. 2.8e-25;
Matches 208; Conservative 135; Mismatches 262; Indels 243; Gaps 33;

OY 60 GTNNYLSYQSHADVDVLAKAQAAMLSVITWGTFTDGLSLDGSVPTIDGNKNGFYQYW 119
DB 55 GTNHACWYRDLDT-----ALRGIRSGM-----NSRVV-----LSNGY---RW 92

OY 120 DPTGAPAYNDGPTGLOGLDYAIASAAHCLRVIVVLND-----W 160
DB 93 TKIPASEVAN-----IISLSRLGFKAILIEVHDTTCYGEDGAACSLAQAVEYW 141

OY 161 KEFGG-MDYQDKWGLPYHDNFYDPTQOAYKNWYHLLNRVNSITGVYKN---DPT 215
DB 142 KEIKSVLDGNEDEVII---NIGNEPYGNNNYQNWYNDTKNAIKALRDAGFKHTIMVDAP 197

OY 216 IFAMELANEPRCVSGTGLTSG-----TCTQATVNVVDQMSAYVKS-IDPNHMYSVGD 268
DB 198 NMGQDSNTRMDNAQIMEADPLNLVFSITHMYGVYNTASKVEEYIKSEFVDKGLPLVIGE 257

OY 269 EGFYIGSTGCGWPNYNDPSDGVNNALLRVK-----NIDECTYHLYPNYWC 314
DB 258 FGHQ--HTDG-----DP-----DEEAIYRAKYKICLFSWSCGNSVYGVLDWVNNWD 305

OY 315 QN--ADWGTQWIKDHIANAAGIGKPT-----PNYDC 388
DB 306 PNNPTPWG-OWYKTNALGTSSTPTPTSTVTPPTPTPTPTPTPTPTPTPTPTPTPTPT 364

OY 339 -----ILEFG-----WOTPRD---SVYQWTQ 359
DB 365 GQIKVLYANKETNSTTNTIRPWLKVNVSGSSSIDLSRVTRIYWTYVDGERAQSAISDWAQ 424

OY 360 TVRTNCEAGWFWMLAGNVGOPY-----PNYDC 388
DB 425 IGASN--VTEKFKVLSVSGGADYILEIGFKSAGQLOPKQDTGEIQMRNKDDWSNYQ 482

OY 389 FN-----VYPSSTATVLAELAIS---TGTSPSPSPSSPSPPSPSPSPASP 436
DB 483 GNDWSIQSMTSYGENEKTAYIDGVLVNGQEPSPGATPAPATPTPTPTPTPTPTPTPT 542

OY 437 ----SASPSPSPSPSPSPSPSPSGG-VKVOYKNNDASGDNOIKGLQLVNTGSSVOL 491
DB 543 TPTVTATPTPTPTPTPTPTPTATGQIKVLYANKETNSTTNTIRPWLKVNVSGSSSIDL 602

OY 492 STVTRYWFTTRDGGSTLVYNCDAAMCGCNIRASFGSVNPAATPTADTYLQLSFTCGT-- 549
DB 603 SRVTIRYWTYVDGERAQSAIS-DWAIGASNVTFKEVKLSVSVSGADYILEIGFKSAGQ 661

OY 550 LAAGSTGEIQNRVKNKSDNSNEDNDYSGTNTAFQDMTKVTVYVNGRLVNGTEPSGT 608
DB 662 LOPGKDTGEIQIRFNKSDNSNYNOGNDWSIQSMTSYGENEKTAYIDGVLVNGQEPSPGT 721

OY 609 SPSPSPSP---SPTPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSP 664
DB 722 TSPSTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPT 781

OY 665 CRATYVNVSDWSGEFTATVTV-TNTGSRTSGWTVWAFSGNOTVTNYNWTALTQSCASV 723
DB 782 SPSPSVVE-----ITINTNAGRTQIS-----PYIYGANQDIEGVVHSARRLGCNRL 827

OY 724 TATNLSYN 731
DB 828 TGYNNWENN 835

RESULT 3
GUNA_CALSA
ID GUNA_CALSA STANDARD: PRT: 1742 AA.
AC P22534;
DT 01-AUG-1991 (Rel. 19, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE Endoglucanase A precursor (EC 3.2.1.4) (Endo-1,4-beta-glucanase A)
```

Query Match	15.3%	Score 632.5	DB 1	Length 1742
Best Local Similarity	22.8%	Pred. No. 4.6e+23		
Matches 227	Conservative 143	Mismatches 290	Indels 335	Gaps
Qy	7	PARAFVTAAGTAAVAAAATLGSITMPSA-----	-TAAPAGFVTASGQGV	50
Db	170	PCSAVVAETAAASAAASIVLKDNPATYLOHAKDLFEFAEVYKSDSCYTAANG-----		225
Qy	51	LNGLPYRYGGTNNYYLSYQSHADV-----	-DDVLAKAQAMNLSVIRTWGFI	96
Db	226	-----YVNSWSGFYDELSWAAVWLYLATNDSTYLTKAE-----SYVQNW-----		264
Qy	97	IGSLDGSVPTIDGNK-----NG-----	-FFYQVWDST	123
Db	265	-----PKISGNIIDYKWAHCWDDVHNGAALLAKITDKDTYKQIIESHLDYTTGY		316
Qy	124	CAPAYNDGPTGLOGLD-----YAIASAAHGLRVIVLTDNWKFEFGGDQYDKWYGLPY		177
Db	317	NGERIKVTPKGLAWLDQWGLRYATTATP--LAFV-----	-YSDMSGCP-	357
Qy	178	IDNFYTDPTQQAQKKNVNHLLNRVNSITGVTYKNDPTIFAWELANPRCVGSGTLPTSG		237
Db	358	-----TGKKETRYKFGEQIDYALGSGCRSP-----	-VVGFGTNPKR	394
Qy	238	TCTQATITVNWVDQMSAVVKSIDPNHNVSVGDEGFYIGSTOGSGWPYNPDSGDVDNALLR		297
Db	395	PHHRTAHSWADS-----QSIPSYRHRTL--YCALVGG--PGSDDSYTDIDISNYVNEVAC		446
Qy	298	VKNTDF-----GTVHLY-----PNYWCQNADWGTOIKDHIANAAIG-----KPTILEEF		343
Db	447	DYNAGFYCALAKMYLLYGGNPIDPFKAIETPTNDEFVFEAGINASCTNFIETKAIANNQS		506
Qy	344	GW-----OTPDSDSVQYQTTQVFTNCEAGNW-----FWMLA		375
Db	507	GWPARATNKLKRYFYVDLSELIKAGYSPNQLTILSTNYNOCAKGYVWDSRNTIYILV		566
Qy	376	GNVNGQYP-----NYDGFNVY-----YPSSTATVILASEALAI--		408
Db	567	DFTCTLIYPGQDKYKKEVGFRIAAPQNVQWDSNDYSFODIKGVSSGSVVKTKYIPLYD		626
Qy	409	-----STCTSP-----PPSPSSPSSPSPSPSPSPSPSPSPSPSPSPSPSPSPS		451
Db	627	EDIKWGEERGTSVSPPTASVTPPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPT		686
Qy	452	SSP-----VSGVKVQYKKNDSAPCDNQIKPGLQVMTGSSVDLSTVTVRY		498
Db	687	STPTVTPPTPVSTPATSCQIKVLYANKETNSTTNTLRPLKVVNSGSSIDLSRVTIRY		746
Qy	499	WFTRDGSSLTLYNCDWAAMCCNIRASFGSVNPATPTADTYLQLSFTGGT--LAAGGST		556
Db	747	WYTVDGERAQSALS-DWAQIGASNVTFKVKLSSSVSGADYYLEIGFKSGAGQLQPKDGT		805
Qy	557	GEIQNRVKNKSDWNFDETNDYSY-GNTAFQDWTKVTVVNGRLVMGTEPESCTSPSPTPS		615
Db	806	GEIQIRNKDDWSNYNOGNDWSIQSTSYGENEKTATIDCVLWVGQEPGCTTPAPTPT		865
Qy	616	P-----SPTPSPSPPTPSPSSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSP		652
Db	866	PTVTVTPPT		925
Qy	653	-----SPSPSVSSSGVGCRTAVVN-----SDWSGGFTATV-----VTNFGS		690
Db	926	IRPLKVVNSGSSIDLSRVTIRYTVYDGERAQSALSDWAQIGASNVTFKVKLSSSVS		985
Qy	691	RATSGMTVAWSFGCQTV-----TNY-----WNITALTOSGASVTAT		726
Db	986	GADYYLEIGFKSGAGQLQPKDGTGEIQIRFNKDDWSNYNOGNDWSIQSTSYGENEKT		1045
Qy	727	NLSY-NNVQPGQSTTFEGNSYSGTNTAPTTLCT		760
Db	1046	--AYIDCVLWVGQEP-----SCATPAPTPTPT		1070

RESULT 4

IND_GUNB_CALSA	STANDARD;	PRT;	1039 AA.
AC	P10474;		
DT	01-JUL-1989 (Rel. 11, Created)		
DT	01-JUL-1989 (Rel. 11, Last sequence update)		
DT	16-OCT-2001 (Rel. 40, Last annotation update)		
DE	Endoglucanase/exoglucanase B precursor [Includes: Endoglucanase		
DE	(EC 3.2.1.4) (Endo-1,4-beta-glucanase) (Cellulase)		
DE	(Cellobiohydrolase); Exoglucanase (EC 3.2.1.91) (Exocellobiohydrolase)		
DE	(1,4-beta-cellobiohydrolase)].		
GN	CELB.		
QS	Caldocellum saccharolyticum (Caldicellulosiruptor saccharolyticus).		
OC	Bacteria; Firmicutes; Clostridia; Clostridiales; Syntrophomonadaceae;		
OC	Caldicellulosiruptor.		
OX	NCBI_TaxID=44001;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RA	MEDLINE=89098398; PubMed=2789517;		
RA	Saul D.J., Williams L.C., Love D.R., Chamley I.W., Berquist P.I.;		
RT	"Nucleotide sequence of a gene from Caldocellum saccharolyticum		
RT	encoding for exocellulase and endocellulase activity.";		
RL	Nucleic Acids Res. 17:439-439;(1989)		
CC	-1- FUNCTION: THIS PROTEIN IS MADE UP OF TWO DOMAINS: THE N-TERMINAL.		
CC	DOMAIN HAS EXOGLUCANASE ACTIVITY WHILE THE C-TERMINAL DOMAIN IS		
CC	AN ENDOGLUCANASE.		
CC	-1- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic		
CC	linkages in cellulose.		
CC	-1- CATALYTIC ACTIVITY: Hydrolysis of 1,4-beta-D-glucosidic linkages		
CC	in cellulose and celotetraose, releasing cellobiose from the non-		
CC	reducing ends of the chains		
CC	-1- SIMILARITY: IN THE N-TERMINAL SECTION; BELONGS TO CELLULASE FAMILY		
CC	F (FAMILY 10 OF GLYCOSYL HYDROLASES).		
CC	-1- SIMILARITY: IN THE C-TERMINAL SECTION; BELONGS TO CELLULASE FAMILY		
CC	A (FAMILY 5 OF GLYCOSYL HYDROLASES).		
CC	-----		
CC	THIS SWISS-PROT entry is copyright. It is produced through a collaboration		
CC	between the Swiss Institute of Bioinformatics and the EMBL Outstation -		
CC	the European Bioinformatics Institute. There are no restrictions on its		
CC	use by non-profit institutions as long as its content is in no way		
CC	modified and this statement is not removed. Usage by and for commercial		
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/		
CC	or send an email to license@isb-sib.ch).		
CC	-----		
DR	EMBL; X13602; CAA31936.1; ..		
DR	PIR; S02711; S02711.		
DR	HSSP; 006851; INBC.		
DR	InterPro; IPR001956; CRD_3.		
DR	InterPro; IPR001547; GH_5.		
DR	InterPro; IPR001000; Glyco_hydro_10.		
DR	Pfam; PF001150; cellulase; 1.		
DR	Pfam; PF00331; Glyco_hydro_10; 1.		
DR	Pfam; PF00942; CBM_3; 1.		
DR	PRINTS; PR00134; GLYHYDRLASE10.		
DR	PRODOM; PD001947; CBD_3; 1.		
DR	PROSITE; PS00591; GLYCOSYL_HYDROL_F10; 1.		
DR	PROSITE; PS00659; GLYCOSYL_HYDROL_F5; 1.		
KW	Cellulose degradation; Hydrolase; Glycosidase; Repeat;		
KW	Multifunctional enzyme; Signal.		
FT	SIGNAL 1 28		
FT	CHAIN 29 1039		
FT	DOMAIN 376 416		
FT	DOMAIN 417 570		
FT	DOMAIN 571 618		
FT	ACT_SITE 177 177		
FT	ACT_SITE 285 285		
FT	ACT_SITE 792 792		
SQ	SEQUENCE 1039 AA: 117641 MW; OE0378171594DDAE CRC64;		

Query Match 14.6%; Score 603.5; DB 1; Length 1039;

Best Local Similarity 28.5%; Pred. No. 6.3e-22;

Matches 179; Conservative 107; Mismatches 209; Indels 133; Gaps 30;

Query Match 14.6%; Score 603.5; DB 1; Length 1039;
Best Local Similarity 28.5%; Pred. NO. 6.3e-22;
Matches 179; Conservative 107; Mismatches 209; Indels 133; Gaps 30;


```

Db 428 QGADTYLELGFKNKGTLPACASTNIQLRHNDMSNVAQSGDYSFFKSNFTKTKTKITLY 487
QY 595 VGNRLVWGTEPS 606
Db 488 DOCKLWGTPEP 499

RESULT 8
GUN1_BACSU
ID GUN1_BACSU STANDARD; PRT; 499 AA.
AC P07983;
DT 01-AUG-1988 (Rel. 08, Created)
DT 01-NOV-1991 (Rel. 20, Last sequence update)
DT 01-OCT-1994 (Rel. 30, Last annotation update)
DE Endoglucanase precursor (EC 3.2.1.4) (Endo-1,4-beta-glucanase)
DE (cellulase).
GS BGLC OR GLD.
ON Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=1423;
RN 11;
RP SEQUENCE FROM N.A.
RC STRAIN=DLG;
RA MEDLINE=87194581; PubMed=3106328;
RX Robson L.M., Chambliss G.H.;
RT "Endo-beta-1,4-glucanase gene of Bacillus subtilis DLG.";
RL J. Bacteriol. 169:2017-2025(1987).
CC -!- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic
links in cellulose.
CC -!- SIMILARITY: BELONGS TO CELLULASE FAMILY A (FAMILY 5 OF GLYCOSYL
HYDROLASES).
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch)
CC
CC EMBL; M16185; AAA22496.1; ALT_INIT.
DR PIR; A26874; A26874.
DR HSSP; 085465; 1A3H.
DR InterPro; IPR001956; CBD_3.
DR InterPro; IPR001547; GH_5.
DR Pfam; PF00150; cellulase; 1.
DR Pfam; PF00942; CBM_3; 1.
DR ProDom; PD001947; CBD_3; 1.
DR PROSITE; PS00659; GLYCOSYL_HYDROL_F5; 1.
KW Cellulose degradation; Hydrolase; Glycosidase; Signal.
FT SIGNAL 1 29
FT CHAIN 30 499
FT ACT_SITE 169 169 PROTON DONOR (BY SIMILARITY).
FT ACT_SITE 237 237 NUCLEOPHILE (BY SIMILARITY).
FT DOMAIN 330 499 CELLULOSE-BINDING (BY SIMILARITY).
SQ SEQUENCE 499 AA; 55187 MW; 339D04EE95A63EE1 CRC64;

Query Match
Best Local Similarity 25.7%; Pred. No. 3.7e-15;
Matches 157; Conservative 93; Mismatches 210; Indels 150; Gaps 27;

QY 22 AAATLGSITMPSATAAGFVTSAGGQFVLNG--LPYRYG-----GTNNYLSYQSHAD 73
Db 15 AVLTMGGLLPSPASAGTKTPVAKNQLSIKGTQLVNRDQKAVQLKGISSHGLWYG--- 71
QY 74 VDDVLAKAAMNLSVIR-TWGEIDIGSLDGSVPTIDGNKNGFYQYWDPSGTGAPYNDGP 132
Db 72 --DFVNKD---SLKWLRLDDWG---ITVFRAAMYTAGD-----GVIDNP 106
QY 133 TGLQGLDYAIAASAAHGLRVIVLTNDWKEFGMDQYDKWGLPYHDNFTDPTQOAYK 192
Db 107 SVKNKVKAEVAEKELGIIVII-----DW-----HILNDGNPNQ-NK 142
```

```

QY 193 NWVNHLLNRVNSITGYTYKNDPTIFAWELANEPKRCVSGTLPTSGTCTOATINWVWQMS 252
Db 143 EKAKEFFKEMSSL-----YGNTPNVI-YEIANEP-----NCD-----VNMKKDKIK 181
QY 253 AYVKSI-----DPNIMVSVGDEGFYIGSTQSGWPNYNDPSDGVNDNALLRVKNIDFGT 305
Db 182 PYAEVIVSVIRKNDPDNIIV-----GTGTHSQDVNDAADDQ--LKDANVWY-A 227
QY 306 YHLYPNYMGONADWGTQWIKDHIANAAAICKPTILEFFGWOTPDSDVSYOTWTOTVITNG 365
Db 228 LHFYAGTHGQS-----LRDKANYALSKGAPFVTE--MGTSQ-----ASPS 265
QY 366 EAG-----WNFMLAGNVNGOPYPNYDGFNVVYPSSTATVVLASEALAISTGTSPPS 417
Db 266 NGVFVLDQSRWLNLYDSKNIS---WVWNLSDKQESSALKKPGASK-----TGGWPLTID 317
QY 418 PSSSPSSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSP 477
Db 318 LTASGTFVRENIRGTRKDTKDPETPAQDNPTQEK-----GVSVOYKAGDGRVNSNQIKP 372
QY 478 GLQLVNTGSSVDLSTVTVRYWF--TRDGGSTLVYNCDMAAMCGNIRASFGSVNPAITPT 536
Db 373 QLHIKNGNATVDLKDVTRARYWYVNVKNKQN---FDCDYAQMGCGNLTHTKFTVLIKPKQG 429
QY 537 ADTYLQLSFTGGTLAAGSGTGEIQNRVKNKSDWSNFDENDYSYCTNTAFODWTKVTVVYN 596
Db 430 ADTYLELGFTKGTLSGPGASTNIQLRHNDMSNVAQSGDYSFFKSNFTKTKTKITLYHQ 489
QY 597 GRLVWGTEPS 606
Db 490 GKLIWGTPEP 499

RESULT 9
GUNW_ERWCA
ID GUNW_ERWCA STANDARD; PRT; 504 AA.
AC Q59395;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE Endoglucanase V1 precursor (EC 3.2.1.4) (Endo-1,4-beta-glucanase V1)
DE (Cellulase V1).
GN CELV1.
OS Erwinia carotovora.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Pectobacterium.
OX NCBI_TaxID=554;
RN 11;
RP SEQUENCE FROM N.A.
RC STRAIN=SCC3193;
RX MEDLINE=95231512; PubMed=7715600;
RA Mae A., Heikinhelmo R., Palva E.T.;
RT "Structure and regulation of the Erwinia carotovora subspecies
RT carotovora SCC3193 cellulase gene celv1 and the role of cellulase in
RT phytopathogenicity.";
RL Mol. Gen. Genet. 247:17-26(1995).
CC -!- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic
links in cellulose.
CC -!- SUBCELLULAR LOCATION: Secreled.
CC -!- SIMILARITY: BELONGS TO CELLULASE FAMILY A (FAMILY 5 OF GLYCOSYL
HYDROLASES).
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch)
CC
CC EMBL; X79241; CAA55823.1;
DR HSSP; 085465; 1A3H.
```



```
DR InterPro: IPR001956; CBD_3.  
DR InterPro: IPR001547; CH_5.  
DR Pfam: PF00150; cellulase: 1.  
DR Pfam: PF00942; CBM_3; 1.  
DR ProDom: PD001947; CBD_3; 1.  
DR PROSITE: PS00659; GLYCOSYL_HYDROL_F5; 1.  
KW Cellulose degradation; Hydrolase; Glycosidase; Signal.  
FT SIGNAL 1 31 POTENTIAL.  
FT CHAIN 32 504 ENDOGLUCANASE V1.  
FT DOMAIN 32 334 CATALYTIC.  
FT DOMAIN 335 352 LINKER.  
FT DOMAIN 353 504 CELLULOSE-BINDING (BY SIMILARITY).  
FT ACT_SITE 168 168 PROTON DONOR (BY SIMILARITY).  
FT ACT_SITE 256 256 NUCLEOPHILE (BY SIMILARITY).  
SQ SEQUENCE 504 AA; 54963 MW; 0D7ECF74781565FA CRC64;  
  
Query Match 10.9%; Score 451; DB 1; Length 504;  
Best Local Similarity 29.0%; Pred. No. 5.1e-15;  
Matches 142; Conservative 64; Mismatches 158; Indels 126; Gaps 24;  
  
QY 143 ASAAHGLRLVLTNDWKEFGGMDQYDKWGLPYHDNEYTDPRTOQAKYKNVHLLNRV 202  
DB 144 AVAAAGLGVYIII--DWHTLS-----DN-----YKAQAKIFFAEM 151  
QY 203 NSITGVVTKNDPTIFAWELANERPCVSGTLPTSGTCTQATIVNVVDMSAY-----V 255  
DB 152 AGL-----YGNSPNVI-VEIANEPN--GS-----VTWNGQIRPYALEVTDTI 190  
QY 256 KSIDPNHMVSGDEGFFVIGTSGCGWPNDSGVDN-----NALLRVKNIDFGTYHLP 310  
DB 191 RSKDPNLIIV-----CGTSWQDIHDAADNOLPDPNTLYAL-----HFYA 231  
QY 311 NYMGONADWCTQWIKHIANAATGKPTILEEFCWQTPDRDS-----VYQTWTQVRT 363  
DB 232 GTHG-----QFLRDIRDAQSRGAIFVSE--WGTSDASGNGGPFLESOTWIDFLNN 282  
QY 364 NGEAGNFWMLAGNVNQPNYDGFVYPSSTATVLAESALISTGTSPPPSPSSPS 423  
DB 283 RG-ISMVNWLSDK-----SETSAALVA-----GASKSGGWTEQNL 317  
QY 424 SSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSP 479  
DB 318 STSCKFVREQIRACAGLSGGDTPTMPTPTNGNCTGTVLQVLRNVNDNPSDDAI RMAF 377  
QY 480 QLVNTGSSVLDLSTVTVRYVTRDG--GSSTLVYVNCDAAMGCNIRASFGSVNPATPTA 537  
DB 378 NIKNTGSTPIKLSDLQVRYVYFHDGKPGANLFV--DMANVGPNPIVSTGTTPAASDTKA 434  
QY 538 DTYLQLSF-TGGTLAAGCGTGEIQNRVNSKDSWNFDETNDYSYGTN-TAFODWTKVTYV 595  
DB 435 NRYVLVTFASGSLQPGAETGEQVRIHAGDSNVNENYDYSYGNITSVTNMDKITVHD 494  
QY 596 NGRVLVWCTEP 605  
DB 495 KGLVWCTEP 504  
  
RESULT 10  
GUND_CELFI  
ID GUND_CELFI STANDARD; PRT; 747 AA.  
AC P50400;  
DT 01-OCT-1996 (Rel. 34, Created)  
DT 01-OCT-1996 (Rel. 34, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Endoglucanase D precursor (EC 3.2.1.4) (Endo-1,4-beta-glucanase)  
DE (Cellulase).  
GN CEND.  
OS Cellulomonas fimi.  
OC Bacteria; Actinobacteria; Actinobacteria (class); Actinobacteridae;  
OC Actinomycetales; Micrococciaceae; Cellulomonadaceae; Cellulomonas.  
OX NCBI_TaxID=1708;  
RN [1]  
RP SEQUENCE FROM N.A.
```


	Query Match	9.6%	Score 394:	DB 1:	Length 145:
	Best Local Similarity	50.7%:	Pred. No. 6.9e-13:		
	Matches	75;	Conservative 31:	Mismatches 36:	Indels 6:
				Gaps 4:	
Qy	461	VQYKNDSAPCDNQIKPGLQLVNTGSSVDLSITVTYYWTFTRDCCSSTI.VYNCDWAAMCIC	520		
Dd	1	LOYRAADTNAADNOIKPFSEVNKNGTSVAVDLSLTKIRYFTFKDGSAAVNCM-IDMAOLGG	54		
Qy	521	GNIRASFGSNPATPTADTVTLQLSFT--GGT'LAAAGSGTGFIQNRYNKSWSNSFETINDYS	578		
Dd	60	SNIOISFG--NHCTGNSDTYVELSFSSFEACSGIAAGCGSETQLRMKSTDMSNFNEANDYS	117		
Qy	579	Y-GRNTAFQDWTKVITYVYVNGRLVMGTPE	605		
Dd	118	FDGTTKTAFAQMDWRVVLYONGIYVHGTAIP	145		

RESULT 15	
GUNA_MICBI	
ID	GUNA_MICBI
STANDARD;	PRT; 456 AA.
P26414;	
DT	01-AUG-1992 (Rel. 23, Created)
DT	01-AUG-1992 (Rel. 23, Last sequence update)
DT	15-DEC-1998 (Rel. 37, Last annotation update)
DE	Endoglucanase A Precursor (EC 3.2.1.4) (Endo-1,4-beta-glucanase)
DE	(Cellulase).
GN	CELA.
OS	Microbispora bisporea.
OC	Bacteria; Actinobacteria; Actinobacteridae (Class); Actinobacteridaceae;
OC	Actinomycetales; Pseudonocardineae; Pseudonocardiaceae;
OC	Thermobispora.
OX	NCBI_TaxID=2006;
RN	[1]
SEQUENCE FROM N.A.	
RA	Yablonsky M.D., Elliston K.O., Eveleigh D.E.;
RL	(In) Coughlan M.P. (eds.);
RL	Production, characterization and application of cellulose,
RL	hemicellulose and lignin enzyme degrading systems, pp.77-83, Elsevier,
RL	London (1989).
CC	-1- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic
CC	linkages in cellulose.
CC	-1- SIMILARITY: CONTAINS 1 BACTERIAL-TYPE CELLULOSE-BINDING DOMAIN
CC	(CBD).
CC	-1- SIMILARITY: BELONGS TO CELLULOSE FAMILY B (FAMILY 6 OF GLYCOSYL
CC	HYDROLASES).
DR	HSPSP; P26222; 1TML.
DR	InterPro: IPR001919; Bac_celose-bind.
DR	InterPro: IPR001524; GH_6.
DR	Pfam: PF00553; CBM_2; 1.
DR	Pfam: PF01341; Glyco_hydro_6; 1.
DR	PRINTS; PR00733; GLHYDRLASE6.
DR	ProDom: PD003733; GH_6; 1.
DR	PROSITE; PS00561; CBD_BACTERIAL; 1.
DR	PROSITE; PS00655; GLYCOSYL_HYDROL_F6_1; 1.
DR	PROSITE; PS00566; GLYCOSYL_HYDROL_F6_2; 1.
KW	Cellulose degradation; Hydrolase; Glycosidase; Signal.
FT	SIGNAL 1 30
FT	POTENTIAL.
FT	CHAIN 31 456
FT	ENDOGLUCANASE A.

FT	DOMAIN	31	322	CATALYTIC.
FT	DOMAIN	323	355	LINKER ("HINGE") (PRO-SER BOX).
FT	DOMAIN	356	456	CELLULOSE-BINDING.
FT	ACT_SITE	113	113	BY SIMILARITY.
FT	ACT_SITE	151	151	PROTON DONOR (BY SIMILARITY).
FT	ACT_SITE	300	300	NUCLEOPHILE (BY SIMILARITY).
FT	DISULFID	114	159	BY SIMILARITY.
FT	DISULFID	267	302	BY SIMILARITY.
FT	DISULFID	360	453	BY SIMILARITY.
SQ	SEQUENCE	456 AA:	47011 MW;	B06D8595E322848F CMC64;
Query Match				
Best Local Similarity 9.2%; Score 380.5; DB 1; Length 456;				
Matches 144; Conservative 63; Mismatches 206; Indels 113; Gaps 24;				
Qy	255	VKSIDPNHWSVGDGCFYIG-STQSGWPNYNDPDCVDNNALLRVKNIDFGTYHLYPNYW	313	
Db	26	IASAGPAIAY---DSPFYVDPOSNAAKVVAANPND--PRTPVIRDIRIAAVPTGRWFANY-	79	
Qy	314	QONADWGQTQIKDHTANAAAGKPTILEEFGWQTPDRDSVYQTWTQTVRTNGEAGWNFMW	373	
Db	80	--NPSTVRAEVDAYVGAAGAAACKIPIMVYA--MPNRD-----	113	
Qy	374	LAGNYNGQYPNYDCGFNYVYPSSTATVLASEALAISTGTPPPSPSSSPSSPSPSPSPS	433	
Db	114	-CGGPSAGAPNITAYRAWIDFIAAGLRNRPAAIIL-----EPDALPIMTNCMSPSEQ	165	
Qy	434	ASPSASPSASSPSPSPSPSPSPSGCGVKQYKNDSA--PGDQNTK--PCQLQVNTGSSSV	489	
Db	166	AEVQASAVGACKKPKAASSQ-----AKVYFDAGHDWVPADEMASRLRGADIANS-ADGI	219	
Qy	490	DLSTVTVRYWFTTRDGSSTLYV-NCDMAAMCGNTRASFQ-SVNPATPTADTYLQLSETG	547	
Db	220	ALNVSNYRY-----TSGLISYAKSVLSAIGASHLRAVIDTSRNGNGPLGSEWCD-----	268	
Qy	548	GTLAAGGSTGEIQNRVKNKSDKSNDETDNDYSYGTNTAFQDWTKVY-----VYVN	596	
Db	269	---PPCRATG-----TWSTTD-TGDPAL---DAFL-WIKPPGEADGCIATPGVFVP	311	
Qy	597	GRLVNCTEPTSPSTPSPSTPSPSTPSPSTPSPSTPSPSTPSPSTPSPSTPSPSPSPSP	656	
Db	312	DR---AYELANNAAPPTYSPSTPS-----TPSPSPSQSDPGSPSPSPSQPP	355	
Qy	657	SVSSGVCGRATYVVVNSDWGSGFTATVTVNTGSRATSGWTVAMSGNGQTVNTYNTAL	716	
Db	356	----AGRACEATYALVNQWPGCFQAEVTVKNTGSSPINQWTVQWTLPSGQSITQLWNGDL	411	
Qy	717	TQSGASVTATNLSYNNVIOPGOSTTFGFGNGSYSGTNTAPTLLTCTAS	762	
Db	412	STSGSNVTVRNVMNGNVNPGAGSTSGFLGSGTG-QLSSSITCSAS	456	

Search completed: November 13, 2002, 11:55:00
Job time : 22.5172 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 13, 2002, 10:50:41 : Search time 70.6164 Seconds
(without alignments)
2223.392 Million cell updates/sec

Title: US-09-917-378-1
Perfect score: 4122
Sequence: 1 MGLVRRPARAFVATAACTAV.....GFNGSYSGTNTAPTTLCTAS 762

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 205047115 residues 671580
Total number of hits satisfying chosen parameters:

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL_21.*
1: sp_archaea.*
2: sp_bacteria.*
3: sp_fungi.*
4: sp_human.*
5: sp_invertebrate.*
6: sp_mammal.*
7: sp_mhc.*
8: sp_organelle.*
9: sp_phage.*
10: sp_plant.*
11: sp_rodent.*
12: sp_virus.*
13: sp_vertebrate.*
14: sp_unclassified.*
15: sp_rvirus.*
16: sp_bacteriap.*
17: sp_archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	841.5	20.4	666	2 Q9RIK7	Q9rik7 thermotoga
2	813	19.7	680	2 Q9RIK9	Q9rik9 thermotoga
3	811	19.7	669	16 Q9X0V4	Q9x0v4 thermotoga
4	790	19.2	694	2 Q9Z187	Q9z187 bacillus st
5	740	18.0	921	2 Q9L8L8	Q9l8l8 caldicellul
6	723.5	17.6	439	3 Q9Z401	Q9z401 agericus bi
7	719.5	17.5	439	3 Q9P893	Q9p893 agericus bi
8	684.5	16.6	930	2 Q9RFK5	Q9rfk5 caldicellul
9	664.5	16.1	437	3 Q99036	Q99036 trichoderma
10	661.5	16.0	1711	2 P96311	P96311 anaerocellul
11	644	15.6	377	3 Q90012	Q90012 aspergillus
12	624	15.1	1751	3 Q9AQ64	Q9aq64 caldicellul
13	619	15.0	578	3 Q96V96	Q96v96 orpinomyces
14	609.5	14.8	1000	2 Q24820	Q24820 thermophili
15	609	14.8	1779	2 Q52374	Q52374 caldicellul
16	602.5	14.6	1770	2 Q9X3P5	Q9x3p5 caldicellul

17	589.5	14.3	1426	2 Q9X3P6	Q9x3p6 caldicellul
18	552	13.4	431	10 Q9FJ23	Q9fj23 arabidopsis
19	544.5	13.2	416	10 Q9FT03	Q9ft03 coffea arab
20	541	13.1	408	10 Q9SG95	Q9sg95 arabidopsis
21	532	12.9	395	10 Q8RVL3	Q8rvl3 lycopersico
22	530	12.9	399	10 Q93WT4	Q93wt4 lycopersico
23	526.5	12.8	411	10 Q9FZ29	Q9fz29 arabidopsis
24	519	12.6	414	10 Q9SG94	Q9sg94 arabidopsis
25	498.5	12.1	996	2 Q9AQH0	Q9aqh0 caldicellul
26	498	12.1	448	10 Q9LZV3	Q9lzv3 arabidopsis
27	498	12.1	468	10 Q8SAY1	Q8say1 oryza sativ
28	495	12.0	431	10 Q9M0H6	Q9m0h6 arabidopsis
29	494	12.0	427	10 Q9FT02	Q9ft02 coffea arab
30	491.5	11.9	439	10 Q8SAY2	Q8say2 oryza sativ
31	491.5	11.9	445	10 Q94J47	Q94j47 oryza sativ
32	489.5	11.9	170	2 Q9RFK6	Q9rfk6 caldicellul
33	466.5	11.3	415	10 Q9FZ03	Q9fz03 lycopersico
34	464	11.3	401	10 Q9FUQ6	Q9fuq6 lycopersico
35	463	11.2	499	2 Q93TJ6	Q93tj6 bacillus su
36	458	11.1	508	2 Q93LD0	Q93ld0 bacillus su
37	451.5	11.0	499	2 Q45532	Q45532 bacillus su
38	450	10.9	499	2 Q52731	Q52731 bacillus sp
39	433	10.5	1915	2 Q9KPL0	Q9kpl0 acetivibrio
40	430.5	10.4	501	2 Q83012	Q83012 bacillus sp
41	427.5	10.4	997	2 Q9Z411	Q9z411 bacillus sp
42	426.5	10.3	369	10 Q48540	Q48540 lycopersico
43	423.5	10.3	442	16 Q9AA10	Q9aa10 caulobacter
44	415.5	10.1	403	10 Q9SKU9	Q9sku9 arabidopsis
45	407	9.9	311	10 Q9LW44	Q9lw44 arabidopsis

ALIGNMENTS

RESULT 1

Q9RIK7	PRELIMINARY;	PRT;	666 AA.
ID	Q9RIK7		
AC	Q9RIK7		
DT	01-MAY-2000 (TREMBLrel. 13, Created)		
DT	01-MAY-2000 (TREMBLrel. 13, Last sequence update)		
DT	01-JUN-2002 (TREMBLrel. 21, Last annotation update)		
DE	Beta-mannosidase (Beta-mannanase).		
GN	MANB OR MAN5.		
OS	Thermotoga neapolitana.		
OC	Bacteria; Thermotogae; Thermotogales; Thermotogaceae; Thermotoga.		
OX	NCBI_TaxID=2337;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=5068;		
RA	Parker K.N., Lam D., Duffaud G., Snead M.A., Mathur E.J., Kelly R.M.;		
RT	"Amino Acid Sequence of beta-mannosidase genes from the		
RT	hyperthermophilic bacteria Thermotoga maritima and Thermotoga		
RL	neapolitana.";		
RN	Submitted (AUG-1998) to the EMBL/GenBank/DDBJ databases.		
RN	[2]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=DSM 5068;		
RX	MEDLINE=21473892; PubMed=11590605;		
RA	Parker K.N., Chhabra S.R., Lam D., Callen W., Duffaud G.D.,		
RT	"Galactomannanases Man2 and Man5 from Thermotoga species: growth		
RT	physiology on galactomannans, gene sequence analysis, and biochemical		
RT	properties of recombinant enzymes";		
RL	Biotechnol. Bioeng. 75:322-333(2001).		
DR	EMBL; Y17981; CAB56856.1; ..		
DR	EMBL; AY033477; AAK53459.1; ..		
DR	InterPro; IPR001547; GH_5		
DR	Pfam; PF00150; cellulase: 1.		
SQ	SEQUENCE 666 AA: 76695 MW: 7EF986115E4C58CD CRC64:		

Query Match 20.4%; Score 841.5; DB 2: Length 666;
Best Local Similarity 43.1%; Pred. No. 6.6e-35;
Matches 175; Conservative 65; Mismatches 123; Indels 43; Gaps 13;

```
QY 29 ITMPSATAAPAGFVTASGGQFVLNGLPYRYGGTNNYLSYQSHADVDVYLAKAQAMNLSV 88
DB 11 VTLPLIVLFANSDFVAVKNGCRFLNGEERFVGSNNYHYKSNRMIDSVLESAGKANGKV 70
QY 89 IRTWCFIDIGSLDGSVPTIDGNKNGFYQYWDPTGTA---PAYNDGPTGLOGLDYATASA 145
DB 71 LRWNGF-----LDGESYCRDN-----TYMHPEPGVFCPLPCTNAQDGFERLDYTVAKA 119
QY 146 AAHLGRVIVLNDWKKEFGMDQYDKYGLPYHDNFTYDPTQOQAYKNNVHLLNRVNSI 205
DB 120 KELGKLLIIVLNNWDDFCGMNQYVRWEGGIHDDFYRNEKIKEEYKVKYVSLINRVNTY 179
QY 206 TCVTYKNDPTIFAWELANPRCVGSGTLPTSGCTCTQATIVNVVDOMSAYVKSIDPNHMS 265
DB 180 TCVPTREPTIMAWELANPRC-----ETDKSGN-----TLVWEVEMSAYIKSLDPNHLVA 231
QY 266 VGDEGF---YIGSTQSG---WPYNDPSDGVNALLRVKNIDFGTYHILYPNYWGQN--- 316
DB 232 VGDEGFNNYEGFRPYGGEAEWAYNGWS-GVDMKKLLEIETVDFCTFHLPSHWGVS PEN 290
QY 317 -ADWCTOWIKDHIAAAGKPTILEEFG--WQTP-DRDSVYQTVTQVVTNGEAGNFW 372
DB 291 YAQWAGAKIEDHIIKAKEGKPVVLEEYGIKPSAPVNRVAIYKLNWDLVYNLGGNGAMFW 350
QY 373 MLAGNVNGQP-----YPNYDGFNVVYPSSTATVLASE-ALAISTG 411
DB 351 MLAGICEGDRDEKGYPDYDGFRIVNDSEAKLIRYAKLFFSTG 396

RESULT 2
QYRIK9 PRELIMINARY: PRT: 680 AA.
AC QYRIK9:
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Beta-mannosidase.
GN MANB.
OS Thermotoga maritima.
OC Bacteria; Thermotogae; Thermotogales; Thermotogaceae; Thermotoga.
OX NCBI_TaxID=2336;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MSR8;
RA Parker K.N., Lam D., Duffaud G., Snead M.A., Mathur E.J., Kelly R.M.;
RT "Amino Acid Sequence of beta-mannosidase genes from the
RT hyperthermophilic bacteria Thermotoga maritima and Thermotoga
RT neopolitana.";
RL Submitted (AUG-1998) to the EMBL/GenBank/DBDJ databases.
DR EMBL: Y17980; CAB56854.1; -.
DR InterPro: IPR004200; Bgal_small_C.
DR InterPro: IPR001547; GH_5.
DR Pfam: PF02930; Bgal_small_C; 1.
DR Pfam: PF00150; cellulase; 1.
SQ SEQUENCE 680 AA; 77687 MW; 5BA3F6982194D1AA CRC64;

Query Match 19.7%; Score 813; DB 2; Length 680;
Best Local Similarity 41.5%; Pred. No. 1.8e-33;
Matches 173; Conservative 65; Mismatches 135; Indels 44; Gaps 13;

QY 19 AVAAATGCSITMPSATAAPAGFVTASGGQFVLNGLPYRYGGTNNYLSYQSHADVDVYL 78
DB 12 SVSAEFLLLIVELSFVLFASDFVAVKNGCRFLNGEERFVGSNNYHYKSNRMIDSVLES 71
QY 79 AKAAWNLNVIRTWGTFIDIGSLDGSVPTIDGNKNGFYQYWDPTGTAAYNDG----PTG 134
DB 72 ESARDMGKVLWNGF-----LDGESYCRDN-----TYMHPEPGVFCVPEGISNAOSG 120
QY 135 LQGLDYATASAAHGLRVIVLITNDWKKEFGMDQYDKYGLPYHDNFTYDPTQOQAYKNN 194
DB 121 FERLDYTVAKAKELGKLVIVLNNWDDFCGMNQYVRWEGGIHDDFYRNEKIKEEYKY 180
```

```
QY 195 VNHLNRVNSITGVYKNDPTIFAWELANPRCVGSGTLPTSGCTCTQATIVNVVDOMSAY 254
DB 181 VSFLNVHNVNTYGVYRREPTIMAWELANPRC-----ETDKSGN-----TLVWEVKEMSSY 232
QY 255 VKSIDPNHMSVGDEGCF---YIGSTQSG---WPYNDPSDGVNALLRVKNIDFGTYHIL 308
DB 233 IKSLDPNHLVAVGDEGFFSYEGFKPYGGEAEWAYNGWS-GVDMKKLLSLETVDFTFHL 291
QY 309 YPNYWGQN-----ADWCTOWIKDHIAAAGKPTILEEFG--WQTP-DRDSVYQTVTQV 361
DB 292 YPSHWGVSPEYTAQWAGAKIEDHIIKAKEGKPVVLEEYGIKPSAPVNRVAIYKLNWDLV 351
QY 362 RTNGEAGNFMWLAGNVNGQP-----YPNYDGFNVVYPSSTATVLASE-ALAISTG 411
DB 352 YDLGGDGAFWLAGICEGDRDEKGYPDYDGFRIVNDSPAEELIRYAKLFFNTG 408

RESULT 3
QYXOV4 PRELIMINARY: PRT: 669 AA.
ID QYXOV4:
AC QYXOV4:
DT 01-NOV-1999 (TREMBLrel. 12, Created)
DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)
DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
DE Endo-1,4-beta-mannosidase.
GN TM1227.
OS Thermotoga maritima.
OC Bacteria; Thermotogae; Thermotogales; Thermotogaceae; Thermotoga.
OX NCBI_TaxID=2336;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MSR8 / DSM 3109;
RX MEDLINE=99287316; Pubmed=10360571;
RA Nelson K.E., Clayton R.A., Gill S.R., Winn M.L., Dodson R.J.,
RA Haft D.H., Hickey E.K., Peterson J.D., Nelson W.C., Katchum K.A.,
RA McDonald L., Utterback T.R., Malek J.A., Linher K.D., Garrett M.M.,
RA Stewart A.M., Cotton M.D., Pratt M.S., Phillips C.A., Richardson D.,
RA Heldelberg J., Sutton G.G., Fleischmann R.D., Eisen J.A., White O.,
RA Salzberg S.L., Smith H.O., Venter J.C., Fraser C.M.;
RT "Evidence for lateral gene transfer between Archaea and Bacteria from
RT genome sequence of Thermotoga maritima.";
RL Nature 399:323-329(1999).
DR EMBL: AE001779; AAD36302.1; -.
DR TIGR: TM1227; -.
DR InterPro: IPR001547; GH_5.
DR Pfam: PF00150; cellulase; 1.
KW Complete proteome.
SQ SEQUENCE 669 AA; 76931 MW; 5659B727305688F5 CRC64;

Query Match 19.7%; Score 811; DB 16; Length 669;
Best Local Similarity 42.8%; Pred. No. 2.3e-33;
Matches 169; Conservative 61; Mismatches 121; Indels 44; Gaps 13;

QY 41 FVTASGGQFVLNGLPYRYGGTNNYLSYQSHADVDVYLAKAQAMNLSVIRTWGTFIDIGSL 100
DB 23 FVKVENGKFKALNGKEPRFTGSGNNYHYKSNRMIDSVLESARDMGKVLRLWNGF----L 77
QY 101 DCSVPTIDGNKNGFYQYWDPTGTAAYNDG----PTGLOGLDYATASAAHGLRVIVL 156
DB 78 DECSYCRDN-----TYMHPEPGVFCVPEGISNAOSGFERLDYTVAKAKELGKLVIL 131
QY 157 TNDWKEFGMDQYDKYGLPYHDNFTYDPTQOQAYKNNVHLLNRVNSITGVYKNDPTI 216
DB 132 VNVWDDFCGMNQYVRWFGTGHDDFYRDEKIKEEYKVKYVSLNVHNTVTGVPYREPTI 191
QY 217 FAWELANPRCVGSGTLPTSGCTQATIVNVVDOMSAYVKSIDPNHMSVGDEGCF---YI 273
DB 192 MAWELANPRC-----ETDKSGN-----TLVWEVKEMSSYIKSLDPNHLVAVGDEGFSNVE 243
QY 274 GSTQSG---WPYNDPSDGVNALLRVKNIDFGTYHILYPNYWGQN-----ADWCTOWIKD 326
DB 244 GPKPYGGEAEWAYNGWS-GVDMKKLLSLETVDFTGTFHLPSHWGVSPEYTAQWAGAKIED 302
```

Qy	327	IIANAAGI	KRPTILEERG	--WQTP-DRDSVYQTWTQTVRTNGEAGWNFWMLAGNVNQGP-	383
				: :	
Db	303	HIIKIAEIGK	PWLEEGYIPKSPAPVNNKTATIRLWNLVDYDLGGDGMFWMLAGIGEGSDR	362	
Qy	383	-----Y	PNYDGFNVYPSSTATVLASE-ALAI	STG	411
		: :	: :	: :	
Db	363	DERGYPDYDQ	GFRIVNDDSPAEALIRYAKLFNTG	397	
RESULT 4					
Q9Z187	1D	Q9Z187	PRELIMINARY;	PRT;	694 AA.
AC	Q9Z187;				
DT	01-MAY-1999	(TReMBUrel. 10, Created)			
DT	01-MAY-1999	(TReMBUrel. 10, Last sequence update)			
DT	01-MAR-2002	(TReMBUrel. 20, Last annotation update)			
DE	Beta-1,4-mannanase.				
GN	MANF.				
OS	Bacillus stearothermophilus.				
OC	Bacteria; Firmicutes; Bacilli; Clostridium group; Bacillales;				
OC	Geobacillus.				
NCBI_TaxID=1422;					
RP	SEQUENCE FROM N.A.				
RC	STRAIN-MCA2184;				
RC	MEDLINE-99013648;	PubMed-9797302;			
RA	Ethier N., Talbot G., Sygusch J.;				
RT	"Gene cloning, DNA sequencing, and expression of thermostable beta-				
KT	mannanase from Bacillus stearothermophilus";				
KL	Appl. Environ. Microbiol. 64:4428-4432(1998).				
DR	EMBL: AF038547; AAC71692.1; -;				
DR	InterPro: IPR001547; GH_5.				
DR	Pfam: PF00150; cellulase; 1.				
DR	SEQUENCE	694 AA; 79734 MW; 262ECABB166A1DB6 CRC64;			
Qy	32	PSATAAPAGFV	TASGGQFVLNGLPYRYGGTNNYLSYQSHADVDVYLAKAOAMNLSVIRT	91	
		:	:	:	
Db	38	PATPTKNEFVYR	KDGLMTCNEKREFVGTNNYLYHKSQMIDVDIESAKMGIVKIL	97	
Qy	92	WGFIDIGSLDGS	VPTIDCNKNGFYFYQWDPSTCAPYNDGP-----TGLQGLDYAI	142	
Db	98	WGFPD-----	GWTSENAHTYMQY-----EMGKYMGECPIKELEGAQNGFERLDYTI	146	
Qy	143	ASAAAHGLRIV	LVLTNDWKEFGMDQYDKWGLPYLDHNFYTDPTQOATYKNVNVHLLNRY	202	
		: :	:	:	
Db	147	YKAKQEGIRLIV	LVLTNNNNFNGMQVYNNIGETNHDLPYTDERIKTAYKNVHYLINRK	206	
Qy	203	NSITGVTPYKND	PTIFANELANPRCVSGTLPPTSGCTQATINWVDONSAYVKSTDPNH	262	
		: :	:	:	
Db	207	NOYTCGIYKNE	PTIMANELANPR-----NDSPTGD-----TLVWDAEMSTYIKSIDPHI	258	
Qy	263	MVSYGDEGCFY	---IGSTQSGGWPNYNDPSDGVNALLRVKNIDFGTYHLYPNYWG---Q	316	
		: :	:	:	
Db	259	LVAVGDEGFFR	SRSSGGFGECGYMYTCYNGVDWDRLLIAKNIDYGFHLYPEHWGISPEN	318	
Qy	317	AD-WGTQWIKD	HITANAACIKPTILEFGWQ----TPDRSVYQTVTQTVRTNGEAGWNFW	372	
		: :	:	:	
Db	319	VEKWGEQYIL	DHLAAGKAKKPVVLEEGYSATSGVGNREMIYDTWNRMTFHEHGGTGAMFW	378	
Qy	373	MLAGNVNQGP	-----YPNYDGFNVYPSSTATVLASEALAINGTCTPSPSPSSPSS	424	
		: :	:	:	
Db	379	LLTG-IDN	PESADENCYYPDYDQGFRIVNDHSSVTNLLTYAKLFNG-----DR	426	
Qy	425	SPSPSPSPSP	SP	484	
Db	427	HVEKEPKVY	EAFPAKP-----QDVRCTYRVKVK-----VASDQHK--VQKVL	467	
Qy	485	GSSSDVLSTV	TVRYWFTTRDGGSGTLLYVNDWAMGCGNTAFSGSVNPTADTYLQLS	544	
		: :	:	:	

[illegible]


```
Db 176 VIAAKANGIRLIVALTNNWADYGGMDVYVYQWVGNGQP-HDLFYNTAATKDAFKSVYRT 234
Oy 198 LLNRVANSITCVTYKNDPTIFANELANPRCVGSGTLPTSGCTQATITVNVMDVSAFYVKS 257
Db 235 FVSR-----YANEPTVMWELANPRCKGS-TCTTSGCTCTTIVTNWAKEMSAFIKT 285
Oy 258 IDPNHNVSGDEGFLGTOGSGWPYNDSDGVNDNALLRVKXIDFGTYHLYPNYWGONA 317
Db 286 IDSNNHLVAIGDEGFL-NQPCAPTYPYQG-SEGVDFAENLAISSVDFAFHSYPEPQGA 343
Oy 318 D---WCTQWIKDHIANAAGIKPTILEEFCWQTPDRDSVYQVWTQVTRTNGEAGWFWML 374
Db 344 DAKAKGTQWITDHAASKRVNKKPVILEEFGVTTNQPDY-VAEWFNEVESGLTGDLIWQA 402
Oy 375 AGNVNGOPVNYDGFVNYVYSSSTATVLALEALAI 408
Db 403 GSHLSTGDTN-DCAVAYPDGVPVYPLMKSHASAM 435

RESULT 8
Q9RFX5
ID Q9RFX5 PRELIMINARY: PRT: 930 AA.
AC Q9RFX5;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Multidomain beta-1,4-mannanase precursor.
GN MANA.
OS Caldicellulosus cellulosovorans.
OC Bacillaria: Firmicutes; Bacillales; Clostridium group; Bacillales;
OC Alicyclobacillaceae; Caldicellulosus.
OX NCBI_TaxID=74586;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20120520; PubMed=10653733;
RA Sunna A., Gibbs M.D., Chin C.W.J., Nelson P.J., Bergquist P.L.;
RT "A gene encoding a novel multidomain beta-1,4-mannanase from
RT Caldicellulosus cellulosovorans and action of the recombinant enzyme on
RT kraft pulp.";
RL Appl. Environ. Microbiol. 66:664-670(2000).
DR EMBL; AF163837; AACF22274.1; .
DR HSP; O06831; INBC.
DR InterPro: IPR001956; CBD_3.
DR InterPro: IPR004302; Chitin_binding_3.
DR InterPro: IPR001547; GH_5.
DR Pfam; PF00942; P_rich_extensin.
DR Pfam; PF00150; cellulase; 1.
DR Pfam; PF03067; Chitin_bind_3; 1.
DR PRINTS; PR01217; PRICHEXTENS.
DR ProDom; PD001947; CBD_3; 2.
DR PROSITE; PS00659; GLYCOSYL_HYDROL_F5; UNKNOWN_1.
KW SIGNAL.
FT SIGNAL.
FT CHAIN 1 33 POTENTIAL.
FT CHAIN 34 930 MULTIDOMAIN BETA-1,4-MANNANASE.
SQ SEQUENCE 930 AA; 101576 MW; 0086638D54D1A2CC CRC64;

Query Match 16.6%; Score 684.5; DB 2; Length 930;
Best Local Similarity 30.9%; Pred. No. 7.7e-27;
Matches 168; Conservative 96; Mismatches 149; Indels 131; Gaps 18;

Oy 304 GYHLYPNYWGQADWGTQWIKDHANAAGIKPTILE-----EFGWOT-----PDRDS 352
Db 143 GTWYLVTRHDQDPTQPLKWSOLEPTFPFQVNTNPPNSSGPDCAEYSQVQLPNKQGRHI 202
Oy 353 VYQWTQTVTRTNGEAGWFWMLAGNVNGOPVNYDGFVNY-----SDSPEAFYNCSDVYFGSGPIAYEFGDPREGGTMI 401
Db 203 IYMIWQR-----SDSPEAFYNCSDVYFGSGPIAYEFGDPREGGTMI 243
Oy 402 ASEALAIAGTSPSPSSPSSPSSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSP 461
Db 244 TEP-----PSGTTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPT 292
```

```
Oy 462 QYKNDSAPGDNOIKPGLQLVNTGSSSVLDLSTVTYRYWFTRDGSGSTLVYNCWMAAGCG 521
Db 293 EYRVGDTSATDNQMKPOLRIVNTGSOAVPLTELKRVYWTKN-STOAEQYFCDMAOIGCS 351
Oy 522 NIRASGCVNPAITADTYLQLSFTGCTLAAGSGTGIEQNRVANKSDNSNDETDNDYSY-C 580
Db 352 NIRAQFVLSQVSGADSYIELSFTGSI PAGGTGIEQNRHIIHFTNMNMYNETDDSYNG 411
Oy 581 TMTAFODWTKVTYVNCRLVMGTPEPSTSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSP 640
Db 412 AQTWGPSTRTILYRNCVLWGTPECGSGSPPTPTPTPT-----PTPTPTPTPTPTPTPT 465
Oy 641 SPSTPTSPSPSPSPSP-----SVSSSGV-----GCRATY 669
Db 466 TSPTPTSGGNLSVNTQGLVGINHPAWRDLRLSSSLQIRSMCANAVRIVLSNGCRWTK 525
Oy 670 VVNSDMGSGFTATVTVTNTGSR-----TSCTVAVWSFGGNOTVTNYW----- 712
Db 526 IPASEVADIISQARTL---CYRAVLEVHDTTGYGEDAAACSMKTAVYNYIELKNVLQAG 582
Oy 713 -NTALTOSGAS-----VTAT-----NLSYNNVIOPCOSTTFCFNGSYSCNT 753
Db 583 ENFVIVNIGNEPYGNVYQNVWTDTRNAVALRAGINNTIMV-DAPNMGDWSFTMKHN 641
Oy 754 APTL 757
Db 642 APTI 645

RESULT 9
Q99036
ID Q99036 PRELIMINARY: PRT: 437 AA.
AC Q99036;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Beta-mannanase precursor.
OS Trichoderma reesei (Hypocrea jecorina).
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC Hypocreales; Hypocreaceae; Hypocrea.
OX NCBI_TaxID=51453;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=RUTC30;
RA Stahlbrand H., Saloheimo A., Velmaannpera J., Penttila M.;
RT "cDNA encoding Trichoderma reesei beta-mannanase.";
RL Submitted (NOV-1993) to the EMBL/GenBank/DBJ databases.
DR EMBL; L25310; AAA34208.1; .
DR HSSP; P00725; IAZ6.
DR InterPro: IPR000254; CBD_fungal.
DR InterPro: IPR001547; GH_5.
DR Pfam; PF00734; CBM_1; 1.
DR Pfam; PF00150; cellulase; 1.
DR ProDom; PD001821; CBD_fungal; 1.
DR SMART; SM00236; fCBD; 1.
KW SIGNAL.
FT SIGNAL.
FT CHAIN 1 19 POTENTIAL.
FT CHAIN 28 437 BETA-MANNANASE.
SQ SEQUENCE 437 AA; 47053 MW; 17513DADE12654A7 CRC64;

Query Match 16.1%; Score 664.5; DB 3; Length 437;
Best Local Similarity 35.3%; Pred. No. 3.4e-26;
Matches 157; Conservative 73; Mismatches 162; Indels 53; Gaps 16;

Oy 18 TAVAAAATLGSITMPSATAAPAGFTVATSGQFVNLGLPYRYGCTNNYYSY-OSHADVDD 76
Db 10 SAATAASAALALQVPVRA--SSFVITISGTQFNIDKVGVFAGTNCYWCSTLNHADVDS 67
Oy 77 VLAKAAMNLVIRTWGFDIGSLDGSVPTIDGNKNGFYQYWDPSGTGAPAYNDGTGLQ 136
Db 68 TFSHTISSSSGLKVVVRWGFNDVNT-----QPSQGIWFOKLS-ATGS-TINTGADGLQ 117
Oy 137 GLDYATASAAHGLRIVVLTNDWKEFGMDQDKWYGLPYHDNFYTDPTQQAAYKNVNV 196
```



```
QY 694 SGWTVAMS-----EGCNOTVTNTYMNALTQSCASVTATNLSYNNV---IQPGQ 738
Db 936 QSVSVASSINPAYIDVKFKVLIKANAGGADYYVEIGFKSGAGVLAAGQSTKEIRLSIQKG- 994
QY 739 STTFGFGSYSGTN 752
Db 995 -----SGSYNQSN 1002

RESULT 13
Q96V96
ID Q96V96 PRELIMINARY: PRT: 578 AA.
AC Q96V96;
DT 01-DEC-2001 (TremBLrel. 19, Created)
DT 01-DEC-2001 (TremBLrel. 19, Last sequence update)
DT 01-JUN-2002 (TremBLrel. 21, Last annotation update)
DE Mannanase ManA.
GN MANA.
OS Orpinomyces sp. PC-2.
OC Eukaryota; Fungi; Chytridiomycota; Neocallimasticales;
OC Neocallimastidaceae; Orpinomyces.
OX NCBI_TaxID=50059;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=PC-2;
RX MEDLINE=21405733; PubMed=11514516;
RA Steenbakkers P.J.M., Li X.L., Ximenes E.A., Arts J.G., Chen H.,
RA Ljungdahl L.G., Op Den Camp H.J.M.;
RT "Noncatalytic docking domains of cellulosomes of anaerobic fungi.";
RL J. Bacteriol. 183:5325-5333(2001).
DR EMBL: AF177206; AAL01213.1; -
DR InterPro: IPR002883; CBD_5.
DR Pfam: PF02013; CBM_10; 2.
SQ SEQUENCE 578 AA; 64425 MW; 482A0F36D36F6F6E CRC64;

Query Match 15.0%; Score 619; DB 3; Length 578;
Best Local Similarity 28.6%; Pred. No. 9e-24;
Matches 173; Conservative 76; Mismatches 203; Indels 152; Gaps 21;

QY 29 ITMPSATAAP--AGVVTAGSGGVFLNGLPYRGCTNNYLSYQSHADVDVLAKQAQMHNL 86
Db 11 LFLTASVAGQFQFQVQDGTNFVVDGCKRYFSGSNTYYLMVSNHERVDLALETYARHLL 70
QY 87 SVIRTWGFTD-----ICSLDGSVPTIDCNKNGFYFYQWDPSTGAPYNDGPTGLQGLD 139
Db 71 NVVRAWAFCECEDATRLVDFSGPEVLNGE-----NWEKVD 107
QY 140 YAIASAAHGLRVIVLTDNWKKEFGMDQYDKWYGLPYHDNYTDPRTQOAYKNVNHLL 199
Db 108 YYLAAAQNRIRVLTNNWTDYGGMDVWVQFGGKYHDEFTNKKDIKGYKQVIKAM1 167
QY 200 NRVNSITGVYKNDPTIFAWELANPRC-VGSGTLPTSGCTQOATIVNKHVDMSAYVKS1 258
Db 168 NRVTYITQLYKDDPTIFSWQLANEARCNGNPHGLPVK-NCNTDITTKMDEIATFIHQE 226
QY 259 DPNHMSVGDGEFYIGSTQSGWPNY-----DPSDGVNALLRVKNIDFGTYHLYPNYWG 314
Db 227 DPNHLVSSGIEG--IGLTPPAGVDKNTVYTYTEGTDYEAISALDSIDYNTVHYMPVGVG 284
QY 315 --QNADKGTQWTKDHIAAAATGKPTILEEFGWQT-----PDRDSV1QTWQTVRTNG 365
Db 285 LKDAYAKDGTWTKAHADYDKKFNKPTVVEWGLSTADNPVLEQRDP1VYTMNNEVLAND 344
QY 366 EAGMH-IWMLAGNVNGQYPNYDGE----- 389
Db 345 LIGWGHMHWI-----VCGEDYIGIGIILDEDEITAVIDPFTKKLYANQCNLDTISIVIT 400
QY 423 -----SSSFPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSP 470
```

```
Db 461 PGAKSTSKNNNTTKTTTTSKASEPTSSSSDECF5IALGFPCCSDNTVTVVSDND--- 517
QY 471 GGNQIKPGIQLQVNTGSSVD-----LSTVTYVYWFTRDGGSSSTLYVNCIMAA 517
Db 518 GDMGVENG-EMCGIGGT1VDNDSCFAKSLGYSCCSCDVVY---TDNDGNMGVENGEW-- 571
QY 518 MCGG 521
Db 572 --CG 573

RESULT 14
O24820
ID O24820 PRELIMINARY: PRT: 1000 AA.
AC O24820;
DT 01-JAN-1998 (TremBLrel. 05, Created)
DT 01-JAN-1998 (TremBLrel. 05, Last sequence update)
DT 01-JUN-2002 (TremBLrel. 21, Last annotation update)
DE Beta-glucanase.
OS thermophilic anaerobe NA10.
OC Bacteria.
OX NCBI_TaxID=67756;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NA10;
RA Miyake K., Machida Y., Hattori K., Iijima S.;
RT "Characterization of a multi-domain cellulase from an extremely
thermophilic anaerobe strain NA10.";
RL Submitted (OCT-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL: AB008029; BAA22939.1; -
DR HSP: O06851; INBC.
DR InterPro: IPR001589; Actbind_actnin.
DR InterPro: IPR001956; CBD_3.
DR InterPro: IPR001547; GH_5.
DR InterPro: IPR001000; Glyco_hydro_10.
DR Pfam: PF00942; CBM_3; 1.
DR Pfam: PF00150; cellulase; 1.
DR PRINTS: PR00134; GLYDRLASE10.
DR ProDom: PD001947; CBD_3; 1.
DR PROSITE: PS00019; ACTININ_1; UNKNOWN_1.
DR PROSITE: PS00591; GLYCOSYL_HYDROL_F10; 1.
DR PROSITE: PS00659; GLYCOSYL_HYDROL_F5; UNKNOWN_1.
SQ SEQUENCE 1000 AA; 113265 MW; B9F659A56A752C6B CRC64;

Query Match 14.8%; Score 609.5; DB 2; Length 1000;
Best Local Similarity 26.8%; Pred. No. 5e-23;
Matches 180; Conservative 116; Mismatches 209; Indels 167; Gaps 31;

QY 156 LTNDWKEFGMDQY--DKWYGI-----PYHDNFTDPRTOOAYKNVNV 196
Db 103 VYNEAIDEGSGDYRRSNWNYICGPEYIEKAFIWAHEADPNKLFYNDYNTENSQKR--Q 160
QY 197 HLLNRVNST--TGVTYKNDPTIFAWELANPRCVSGTLPTSGCTQOATI-VNW--VDM 251
Db 161 FIYNLKSLEKGV-----PIHGVGLQCHINDWPSISEI 195
QY 252 SAYVK--SIDPN---HMVSGDEGYIGSTQSGWPNYNDPSCVDNALLRVKNIDFGTY 306
Db 196 ENTIKLFSIPGIEITHTEL-DMSFY---QWGSSTNYAVPPRELLIKOARYKEL-FDLF 250
QY 307 HLYPN-----YWGONADGTOWIKDHTANAAAIK---PTILEEFGWQTPDRDSVYQW 357
Db 251 KKKYKVVNTVTFWGLKDDY--SWLSKN-----YCKDYPLLFDE----- 287
QY 358 TOTVRTNGEAGNFWMLAGNVNGQYPNYDGFNVYPS---STATVLASEALAISTGTSP 414
Db 288 -----NYMSKYAFNSLID-----PSIVPTTTLTPPEIOLTLPT 324
QY 415 P-----PSPSSSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSP 468
Db 325 PTASVTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPT 384
```

QY	469	APCDNOIKPGLQLVNTGSSVDLSTVTVRYWFTRDGSSFLVYVNCDAAMGCGNIRASFG	528
Db	385	NSTTNTIRPLKVVNTGSSIDLSTVRYWFTVDDKAQSAVS-DWAQIGASNTVTKFV	443
QY	529	SVNPATPTADTYLQLSFTCGT--LAAGSTGEIQNRVKNKSDMSNFDETNDYSY-GTNTAF	585
Db	444	KLSSSVSGADYYLEIGFKSGAGLQPKDGTGETQIRFNKSDWSNYNOGNDWSIQSMTSY	503
QY	586	ODMTKVTVVNGRLVNGTEPSC-----TSPSTPSPSTPSPSTPSPSTPSPSTPSPSPSP	639
Db	504	GENMKVTATIDGVLVNGQEPGTATAPIATPTPTAPTATPTPTPTPTPTPTPTPTPTPT	563
QY	640	PSPTSP	682
Db	564	PTITATPAPTAPTSTPAYLDNDNDLVSCNKIVDKDG---RPWLTVGNW-FCY---	616
QY	683	VTVNTGSRATSGWTVNWSFGGNOTVNTWNTALTQSGASVTA-TNLSYNNVIOQSQSTT	741
Db	617	----NTGTVNFDG---VWSCNLKSTLAETANRGNLRLVPISAELILNWSOGIYKPKPIN	669
QY	742	FGFNGSYSCNTT 753	
Db	670	YYVNPKECKNS 681	
RESULT 15			
ID	O52374	PRELIMINARY; PRT: 1779 AA.	
AC	O52374	01-JUN-1998 (TremBLrel. 06, Created)	
DC	01-JUN-1998	(TremBLrel. 06, Last sequence update)	
DT	01-JUN-2002	(TremBLrel. 21, Last annotation update)	
DE	Family 10	xylanase (EC 3.2.1.8).	
GN	XYNC		
OS	Caldicellulosiruptor sp. R69B.1.		
OC	Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridia;		
OC	Clostridiales; Syntrophomonadaceae; Caldicellulosiruptor.		
OX	NCBI_TaxID=70295;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN-RT69B.1;		
RA	Morris D.B., Gibbs M.D., Ford M., Thomas J., Bergquist P.L.;		
RT	"Family 10 and 11 xylanase genes from Caldicellulosiruptor sp.		
RT	R69B.1.";		
RL	Submitted (DEC-1997) to the EMBL/GenBank/DBJ databases.		
DR	EMBL: AF036924; AAB95326.1; --		
DR	HSSP: Q06851; INBC		
DR	InterPro: IPR001956; CBD_3.		
DR	InterPro: IPR005084; CBM_6.		
DR	InterPro: IPR003305; CBM_CenC.		
DR	InterPro: IPR001000; Glyco_hydro_10.		
DR	Pfam: PF00942; CBM_3; 3.		
DR	Pfam: PF02018; CBM_4; 9; 2.		
DR	Pfam: PF03422; CBM_6; 1.		
DR	Pfam: PF00331; Glyco_hydro_10; 1.		
DR	PRINTS: PR00134; GLHYDRLASE10.		
DR	PRODom: PD001947; CBD_3; 3.		
DR	PROSITE: PS00591; GLYCOSYL_HYDROL_F10; 1.		
KW	Glycosidase; Hydrolase; Xylan degradation.		
SQ	SEQUENCE 1779 AA; 194304 MW; CE5269B6806B5CED CRC64;		
Query Match 14.8%; Score 609; DB 2; Length 1779;			
Best Local Similarity 25.9%; Pred. No. 9.9e-23;			
Matches 189; Conservative 116; Mismatches 213; Indels 212; Gaps 28;			
QY	25	TLGSTIMPSATAAPAG--FVTASGQGFVLNGLPYRYGGTNNYYLSQSHADVDDVLAKAQ	82
Db	702	TVSATPTTAPTASPAGSGYWTPTSES-----YGALKVWYANGNMSSTTNVLNPKIK	751
QY	83	AMNLSV-----IRTWGFDID-----IGS-----LDGSVPTIDGNKNG--FYFQ	117
Db	752	ITENGTAVDLNRVKRVRYWYTIIDGAAOSVSVASSINPAYIDRVVVKLGANAGGADITYE	811

QY	118	YWDPSS-----TCAPAYNDGPTCUGLDYALASAAAHLGRVIVVLTN	158
Db	812	VFKSGAGVLAACQSTKEIRLSIQKSSGSYN-----OSNDYSVRSANSY-----IN	858
QY	159	DMKEFGMDQYDKWYGLPYHDNFTYDPTQOQAYKKNVNHLLNRVNSITGVYTKNDPTIFA	218
Db	859	E-KVTCYIDDLVW-----GREPGRNAQIKVW--YANGNLSSTPTNVL--NPKI--	901
QY	219	WELANPRCVGSGTLTTS-----CTCQOATIVNVWDOMSAYVKSIDPNH-----M	263
Db	902	-KIEN-----VGTAVDLNRVKRYWYTIIDGATQSVS-----ASSINPAYIDRVV	947
QY	264	VSVG-----DEGFYIG-----STGSGMPYNDPSPGVNNALLRV	298
Db	948	VKLGANAGGADYYVEVGFKSGAGVLAAGOSTKEIRLSIQKSSGSYNOSNDYSVRSANSY	1007
QY	299	KNIDFGTYHILYPNYWQGNADMGTWIKDHITANAAALCKPTILEEFGWQTPDRDSVYQWT	358
Db	1008	ENEKVTGYIDDLVWCKEPSRGT-----KP-----	1032
QY	359	QIVRTNGEAGNFWMLAGNVNCOPIYNYDGFNVVYSSSTATVLASEALAISTGTSHIPPSP	418
Db	1033	-----AGEVTPAPTPT-----STPTPTPTTAPTAPTAPTAPTAPTAPTAPTAPT	1072
QY	419	SSSPSSSP	477
Db	1073	TPPTAPT	1132
QY	478	GLQLVNTGSSVDLSTVTVRYWFTRDGSSFLVYVNCDAAMGCGNIRASFGSYNPATPTA	537
Db	1133	WFKVINGGSSVDLSRVKIRYWTVDGDKPQSAV-CDMAQIGASNVTFFVKLTSGVSGA	1191
QY	538	DYVLQLSFTTGGT--LAAGSTGEIQNRVKNKSDMSNFDETNDYS-----YCTNTAFQDW	588
Db	1192	DYILEYGFSSGAGLOPKDGTGDIQVRFKNKNDWSNTNOADDWSNMOSMTNYGEN-----	1245
QY	589	TKTVTVVNGRLVNGTEPSCSTPSPSTPSPSTPSPSTPSPSTPSPSPSPSPSPSPSPSP	648
Db	1246	AKVTLYVDGVLVNGQEPGATAPTATPTPTPTATPTPTPTPTPTPTPTPTPTPTPTPT	1305
QY	649	SPSPSPSPSV 658	
Db	1306	TPPTPTPTKV 1315	

Search completed: November 13, 2002, 11:54:07
Job time : 75.6164 secs

Db 12 SVSAEFLLLIVELSVLFASDEFVKVENGKFAALNGKEFRFGSNNYMHYKSNMGIDSVL 71
Qy 79 AKAQANLNSVIRTWGFDIGSLDGSVPTTDGKNKGFFYFOYWDPTSTCAPAYNDG----PTG 134
Db 72 ESARDMGIKVLRWGR-----LDGESYCRDKN-----TYMHPEPGVFGVPEGTSTNAQSG 120
Qy 135 LOCLDYAIAASAAHGLURVIVLTNDWKEFGGMDYDKWGLPYHDNFYTDPRTOQAYKNN 194
Db 121 FERLDYTVAKAKELGKLVILVNNWDDFGGMNQYVRWFGGTHDDFYRDEKIKEEYKKY 180
Qy 195 VNNHLLNRVNSITGVYTKNDPTIFAWELANEPKRCVSGTLPTSGTCTQATIVNWDQMSAY 254
Db 181 VSLFVNHVNTYGVVREPTIMAWELANEPKRC-----ETDKSGN-----TLVEWVKEMSSY 232
Qy 255 VASIDPNHMSVSGDECF---YIGSTCGSG---WPYNDPSDGVNNDLRLVKNIIDFGTYHL 308
Db 233 IKSLDPNHLVAVGDEGFFSNYEGFKPYGGEAEWAYNGWS-GVDWKKLLSIETVDFGTGHL 291
Qy 309 YPNYWGON----ADWCTOMIKOHIANAAAIKPTILEEFG--WQTP-DRDSVYQVWTQV 361
Db 292 YPSHMCVSPENYAQWCAKAKWIDHIIKIAKEIGKPVVLEEYGIKPSAPVNRRTAIYRLANDLV 351
Qy 362 KTNCEAGHFWMLAGNVNQP-----YPNYDGFNVYPSSTATVLAASE-ALAI 411
Db 352 YDLGGDCAMFWMLAGIGEGSDRDERGYYPDYDGFRIVNDSDPEAEIIRYAKLFNTG 408

RESULT 2
US-09-134-078-61
: Sequence 61, Application US/09134078
: Patent No. 6368844
: GENERAL INFORMATION:
: APPLICANT: Bylina, Edward J.
: TITLE OF INVENTION: GLYCOSIDASE ENZYMES
: NUMBER OF SEQUENCES: 72
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Gray Cary Ware & Freidenrich LLP
: STREET: 4365 Executive Drive, Suite 1500
: CITY: San Diego
: STATE: CA
: COUNTRY: USA
: ZIP: 92121
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Diskette
: COMPUTER: IBM Compatible
: OPERATING SYSTEM: Windows95
: SOFTWARE: FastSeq for Windows Version 2.0
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/09/134,078
: FILING DATE: 13-AUG-1998
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: 08/949,026
: FILING DATE: 10-OCT-1997
: APPLICATION NUMBER: 60/056,916
: FILING DATE: 06-DEC-1996
: ATTORNEY/AGENT INFORMATION:
: NAME: Haile, Lisa A.
: REGISTRATION NUMBER: 38,347
: REFERENCE/DOCKET NUMBER: 09010/024002
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 858/677-1456
: TELEFAX: 858/677-1465
: INFORMATION FOR SEQ ID NO: 61:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 663 amino acids
: TYPE: amino acid
: TOPOLOGY: linear
: MOLECULE TYPE: protein
: "PAGE" TYPE: Internal
ns-0 134-078-61

Query Match 19.7%; Score 812; DB 4; Length 663;
Best Local Similarity 42.8%; Pred. No. 6e-48;
Matches 169; Conservative 61; Mismatches 121; Indels 44; Gaps 13:
Qy 41 FVTASGGQFVLNGLPYRYGCTNNYLSYOSHADVDDVLAKAQAANLNSVIRTWGFDIGSL 100
Db 17 FVKVENGKFAALNGKEFRFGSNNYMHYKSNMGIDSVLESARDMGIKVLRWGR-----I 71
Qy 101 DGSVPTTDGKNKGFFYFOYWDPTSTCAPAYNDG---PTGLOGLDYAIASAAHGLKVLVIL 156
Db 72 DGESYCRDKN-----TYMHPEPGVFGVPEGTSTNAQSGFERLDYTVAKAKELGKIKVIL 125
Qy 157 TNDWKEFGGMDYDKWGLPYHDNFYTDPRTOQAYKNNVHLLNRVNSITGVYTKNDPT 216
Db 126 VNNWDDFGGMNQYVRWFGGTHDDFYRDEKIKEEYKKYVSLFVNHVNTYGVVREPT 185
Qy 217 FAWELANEPKRCVSGTLPTSGTCTQATIVNWDQMSAYKNSIDPNHMSVSGDEGF---Y 273
Db 186 MAWELANEPKRC-----ETDKSGN-----TLVEWVKEMSSYIKSLDPNHLVAVGDEGFFSNYE 237
Qy 274 GSTQSGS---WPYNDPSDGVNNDLRLVKNIIDFGTYHILYPNYWGON----ADWCTOMIKD 326
Db 238 GFKPYGGEAEWAYNGWS-GVDWKKLLSIETVDFGTGTHILYPSHMCVSPENYAQWCAK 296
Qy 327 HTANAAAIKPTILEEFG--WQTP-DRDSVYQVWTQVTRTNGEAGHFWMLAGNVNQP- 382
Db 297 HIKIAKEIGKPVVLEEYGIKPSAPVNRRTAIYRLNDLVYDLGGDCAMFWMLAGIGES 356
Qy 383 -----YPNYDGFNVYPSSTATVLAASE-ALAI 411
Db 357 DERGYYPDYDGFRIVNDSDPEAEIIRYAKLFNTG 391

RESULT 3
US-08-276-213-3
: Sequence 3, Application US/08276213
: Patent No. 5536655
: GENERAL INFORMATION:
: APPLICANT: Thomas, Steven
: APPLICANT: Laymon, Robert
: APPLICANT: Himmel, Michael
: TITLE OF INVENTION: GENE ENCODING FOR THE E1 ENDOGLUCANASE
: NUMBER OF SEQUENCES: 6
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: National Renewable Energy Laboratory
: STREET: 1617 Cole Boulevard
: CITY: Golden
: STATE: CO
: COUNTRY: USA
: ZIP: 80401-3393
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/276,213
: FILING DATE:
: CLASSIFICATION: 435
: ATTORNEY/AGENT INFORMATION:
: NAME: O'Connor, Edna
: REGISTRATION NUMBER: 29,252
: REFERENCE/DOCKET NUMBER: NREL IR# 94-08
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (303)231-1000
: TELEFAX: (303)231-1098
: INFORMATION FOR SEQ ID NO: 3:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 521 amino acids
: TYPE: amino acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: protein

CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/932,571
FILING DATE: September 19, 1997
ATTORNEY/AGENT INFORMATION:
NAME: Bak, Mary E.
REGISTRATION NUMBER: 31,215
REFERENCE/DOCKET NUMBER: 1997US001/C1P
TELECOMMUNICATION INFORMATION:
TELEPHONE: 215-540-9200
TELEFAX: 215-540-5818
TELEX: <Unknown>
INFORMATION FOR SEQ ID NO: 43:
SEQUENCE CHARACTERISTICS:
LENGTH: 1426 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: No. 6294366e
SEQUENCE DESCRIPTION: SEQ ID NO: 43:
US-09-136-574A-43

Query Match 14.3%; Score 589.5; DB 4; Length 1426;
Best Local Similarity 24.6%; Pred. No. 3.3e-32;
Matches 199; Conservative 138; Mismatches 251; Indels 221; Gaps 33:

Qy 84 MHLVSIVRTWGFIDF---GSLGDSVPTI-DGKNKGFYFQWDPSP-----T 123
Db 19 MSISLVGMSYFPVTEAAPDWSIFSLWESYKNDKFGIVAIPARCLSNDDTKQMWLKIKN 78
Qy 124 GAPAYND-----GPTGLOGLOVAIASAAH-----GLRVIVVL-----TNDWKE 162
Db 79 SITAENKMPESLLACQTS-TGLSYRFSTADTFVFANFTNNIGIRGHTLVVHNQTPDW-- 135
Qy 163 FCGMDQYDKWYGLPYHDFNYDPTQQAYKKNVHLLARVNS-ITGVYKKNPTIFAWEL 221
Db 136 -----FFRDSQQMLSK---DALLARKKYIYDVVGRYKGVYAMDV 174
Qy 222 ANEPRCVSGTLPTSGCTQATIVNVQDMSAYKVSIDPNHMSVGDGCFYIG----- 274
Db 175 VNEAIDESQPCYRRSTWYQICGPEY IEKAFIWAHEADPNAKLFYNDYTELSTKRDFIY 234
Qy 275 -----STQSGWP-----YNDPSGDVDNNAL-----LRVKNIDFCTYHLY 309
Db 235 NMVKNLKSQGVPIHGIGMQSHINWMPVSSEIENSIKLFSSIPGIEIHTELDMSLYN-- 292
Qy 310 PNYWCONADMGTOWKDHIANAAGKGP--TILEFEG-----WQTPDRDSVYQWTQ 359
Db 293 ---YGSNENYSTP-PQDLLQROAQYKIDFTWLRKYKGIYCVTWGLKDDOYS----- 341
Qy 360 TVRTNGEAGWFMLAGHNVNGOPYPNYDGFNVYYPSPSTATVLASPA---CTGTSPPPSPS 419
Db 312 -----WLNSSSKRDWPLFFEDDYSA-KPAYWSVIEAACASA-----SPS 379
Qy 420 SSPSSSP 478
Db 380 PNYATPT 433
Qy 478 VYVHTGSSVLDLSTVTVYV 538
Db 434 PKIVNGGSSVLDLSTVTVY 492
Qy 539 TVIOLSTFGGT--LAAGSGTGEIONRVKNSDMNFDENYS-----YGTNTAFQDWT 589
Db 493 YYLEVGFSSGAGOLQPGKRDAGDIQVFNKNKDNWSYVQADDSWLSQSMYDGEN----- 546
Qy 590 KVTYVNGRLVWGTGCTSP 649
Db 547 KVTLYDGVLVWQEPGATPATATATPTPIPTATVTPPTATPTSTPRPTATATPTPT 606
Qy 650 PSP 690
Db 607 VSATPTAPTASPVG--GSYWTSPSEYALKVWYANGNLSSTFVNLNPKIKIENVGTTAV 664

Qy 691 -----RATSGWTVAWS-----EGNQTVTYNYWNTALTQSGASVTAT 726
Db 665 DLSRVKRVYWTIDGATOSVSVASSINPAYIDVKILGANAGGADYVVEIGFKSGAGVLA 724
Qy 727 NLSYNNV---IQPGQSTTFGFGNGSYSGTN 752
Db 725 GOSTREIRLSIQKG-----SGSYNOSN 746

RESULT 7
US-09-033-537A-1
; Sequence 1, Application US/09033537A
; Patent No. 5958083
; GENERAL INFORMATION:
; APPLICANT: Onishi, Masahiro
; APPLICANT: Fich, Merete
; APPLICANT: Toft, Annette Hanne
; APPLICANT: Sh lein, Martin
; TITLE OF INVENTION: Prevention Of Back-Staining
; TITLE OF INVENTION: In Stone Washing
; NUMBER OF SEQUENCES: 1
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 5958083o No. 5958083disk of No. 5958083th America, Inc.
; STREET: 405 Lexington Avenue
; CITY: New York
; STATE: NY
; COUNTRY: U.S.A.
; ZIP: 10174
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/033.537A
; FILING DATE: 02-MAR-1998
; CLASSIFICATION: 008
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 0993/95
; FILING DATE: 08-SEP-1995
; APPLICATION NUMBER: PCT/DK96/00364
; FILING DATE: 03-SEP-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Green, Reza
; REGISTRATION NUMBER: 38,475
; REFERENCE/DOCKET NUMBER: 4492.204-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-867-0123
; TELEFAX: 212-878-9655
; TELEX:
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 551 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-09-033-537A-1

Query Match 11.8%; Score 488; DB 2; Length 551;
Best Local Similarity 25.2%; Pred. No. 8.8e-26;
Matches 157; Conservative 86; Mismatches 165; Indels 216; Gaps 25:

Qy 125 APAYNDGPTGLOGLDYA-----IASAAHGLRVIVVLTNDWKEFGMDQYDKWYG-- 174
Db 1 APAPVFGOLKVQGNOLVGQSQOAVQLVGNSSHGL-----OWYGNF 40
Qy 175 -----LPVHDNFYTDPTQQAYKKNVN-----HLL 199
Db 41 VNKSLQHWNRDNGINVFRAAMYTAEDGYIDPSVKNKVKEAQASIDLGLVVIDWHIL 100
Qy 200 NRVSNT-----GVTYKNDPTIFAWELANEPKRCVSGTLPTSGCTQATIVNW 247


```

: APPLICATION NUMBER: PCT/DK91/00013
: FILING DATE: 18-JAN-1991
: ATTORNEY/AGENT INFORMATION:
: NAME: Zelson, Steve T. / Lambiris, Elias J.
: REGISTRATION NUMBER: 30,335 / 33,728
: REFERENCE/DOCKET NUMBER: 30,335 / 33,728
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 212 867 0123
: TELEFAX: 212 867 0298
: INFORMATION FOR SEQ ID NO: 2:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 700 amino acids
: TYPE: AMINO ACID
: TOPOLOGY: linear
: MOLECULE TYPE: protein
: US-07-862-588B-2

Query Match      10.68; Score 438; DB 2; Length 700;
Best Local Similarity 25.08; Pred. No. 3.3e-22;
Matches 187; Conservative 109; Mismatches 253; Indels 200; Gaps 34;

QY 5 RPARAFVATAAGTAVAAATLGSIIMPSTATAAPAGFVTASGGOFVLNGLPYRYGGTNNY 64
DB 4 RQRRLRFVS-----AALAVSLTM-TVPMPSVNAASDVT-----FTTQSERAAISPNI 53
QY 65 YLSYQSHADVDVLAKAQAAMN-----LSVIRTWGFDIGSL--DGSVPTIDGNK 111
DB 54 YGTNODLSGTEWSSRLGGNRLTGYWNENASSAGRDWLHYSDDFLCGNGGVPTDCDK 113
QY 112 NGFYQWIDSTGAPAYNDGPTGLOGLDYATASAAHGLRVIVLTDWKEFGMDQYD- 170
DB 114 -----PGAVVTAFHD--KSLENGAYSIVT-----LQMGAYVSRD--KNGPVDESET 155
QY 171 ----KWYGLPYHDN----FYDTPRTQAYKN-WYNHLLNRVNSITGVYTKNDPT---IFA 218
DB 156 APSPRMKDVEFAKNAPFSLOPDLNDGQVYWDVEYNFLVNR-----YGNASTSTGIKA 207
QY 219 WELANERPCVSGTLPT--SGTCQATIVNVVDMSAYVKSIDPNHMV---SVGDEGYI 273
DB 208 YSLDNEP-ALWSEHPRIHPHEQLOAAELVAKSIDLSRAVKNVDPAETFGPALYFGGAYL 266
QY 274 GSTOGSGWP-----YNDPSDGV---DNNALLRVKNIDFGYHYLYPNYWG---- 314
DB 267 SLQADPWPSLOGNYSWFIDYIYLDQMKNAHTQNGKRLDVLVD-----HWYPEAQGGGQR 321
QY 315 -----QNADWGTQWIKDHIANAANA--I-----GKPT 338
DB 322 IVFGGAGNIDTKARVQAPRSLWDPAQEDSDWIGTWFFSSYLPLIPKLOSSITQYYPGTL 381
QY 339 ILEEF-----GHQTPDRDSVYQTTOTVTRTNGEAGNFWMLAGNVN-----GOPY 383
DB 382 AITEFSYGGDNHISGGIATADALGIFGKY-----GVYAANYWQTEDNTDYTSAAYKLY 434
QY 384 PNVDGNNVYVPS-----STATYLAS-----EALA--TSPPSPSSSPS- 423
DB 435 RNYDGNKSGGSGIKVDAATSDTENSIVYASVTDENSELHLIVLUNKHDDPINATFOLSG 494
QY 424 -----SSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSP 460
DB 495 DKTYTSGRVNGFDQGTSDITEQAATINNNQFTYTLPLLSAYHVLK...S...EPVNSDLV 554
QY 461 VOYKNND-SAPGNOIKPGLQLVNTGSSVDLSVTIVRYWFTRGDGGSTLVYNCDDWAMG 519
DB 555 VOYKDGDRNNATDQIKPHFNIONKQSPDLSSLTLYRYFTKD-SSAAHNGWIDWAKLG 613
QY 520 CGNIRASFGSVNPATPTADTYLQLSFT--GGTLAAGSGTGEIONRVNKSQWSNFDETNDY 577
DB 614 GSNLIQISFGNHGA--DSDTYAELGFSGGAGSAGGQSGQSGEIQLRMSKADWSNPFNEANDY 671
QY 578 SY-GTNTAFODTKVTIVYVNGRLVWGTPEP 605
DB 672 SFDCAKTAYIDWRVTLYQDQQLVWGTPEP 700
```

RESULT 10

```

US-09-198-956-10
: Sequence 10, Application US/09198956
: Patent No. 6165769
: GENERAL INFORMATION:
: APPLICANT: Andersen, Lene N.
: APPLICANT: Schulein, Martin
: APPLICANT: Lange, Niels Erik K.
: APPLICANT: Bjornvad, Mads E.
: APPLICANT: Schnorr, Kirk
: TITLE OF INVENTION: Pectin Degrading Enzymes From Bacillus
: FILE OF INVENTION: Licheniformis
: FILE REFERENCE: 5377-200-US
: CURRENT APPLICATION NUMBER: US/09/198,956
: CURRENT FILING DATE: 1998-11-24
: EARLIER APPLICATION NUMBER: 1344/97
: EARLIER FILING DATE: 1997-11-24
: EARLIER APPLICATION NUMBER: 60/067,240
: EARLIER FILING DATE: 1997-12-02
: NUMBER OF SEQ ID NOS: 26
: SOFTWARE: FASTSEQ for Windows Version 3.0
: SEQ ID NO 10
: LENGTH: 493
: TYPE: PRT
: ORGANISM: Bacillus licheniformis
US-09-198-956-10
```

Query Match 10.18; Score 415.5; DB 4; Length 493;
Best Local Similarity 27.38; Pred. No. 7.3e-21;
Matches 112; Conservative 71; Mismatches 126; Indels 101; Gaps 14;

```

QY 247 WVDQMSAYVKSIDPNHMVSVGDEGYIG-----STOGSGWPY-----NDUS 287
DB 135 WVDHNELY-----HSLNV-DKDYDGLDFDKVDRDAEYITFSWVYHDGWSMLMGSSDS 186
QY 288 DGVD-----NNALLRVKNIDFGYHYLYPNYWGQNAQWGTOWIKDHIANAANAICKP 337
DB 187 DYNRTITTFHHWFENLSRVSPFRGEGHYNYFNKIIDSG----- 229
QY 338 TILFEGHQTDRDSVYQTTOTVTRTNGEAGNFWMLAGN--VNGO----PYPNYDGFNVY 392
DB 230 -INSRMGARIRIENLFENAKDPIVSWYSSSPGYHVSNKFNVSRCGSMPTTSTTTYNPP 288
QY 393 YPSSTATV----LASEALAISTCTSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPS 447
DB 289 YSISLDNVNVKSVKONAGVGKIQRRPPTP--TPTSPS----- 326
QY 448 PSPSSPVSGGVKVOYKNNDAPGDNOIKPGLQLVNTGSSVDLSVTIVRYWFTRDGSS 507
DB 327 ---ANTPVSGNLKVEFYNSNPSTTNSINPOFKVTNTGSSAIDLSKLTLYRYTYVDCOKD 383
QY 508 TLVYNCOWAAM-----CCGNIRASFGSVNPATPTADTYLQLSFTGGTLAAGSGTGE 558
DB 384 QTEW-CDHAAIIGSGSYNGITSNVKGTFTVKMSSSTNNADTYLEISFTGTLPCAHV-Q 441
QY 559 IONRVNKSQWSNFDETNDYSYGTNTAFODTKVTIVYVNGRLVWGTPEPST 608
DB 442 IQGRFAKNDMSNYTQSDNDSFKRSQFVEMDOVTAYLNGVLVWKGPEGGS 491
```

RESULT 11

```

US-09-198-955A-12
: Sequence 12, Application US/09198955A
: Patent No. 6187580
: GENERAL INFORMATION:
: APPLICANT: Andersen, Lene N.
: APPLICANT: Schulein, Martin
: APPLICANT: Lange, Niels E.
: APPLICANT: Bjornvad, Mads E.
: APPLICANT: Moller, Soren
: APPLICANT: Glad, Sanne O. S.
: APPLICANT: Kauppinen, Markus S.
```

APPLICANT: Schnorr, Kirk
APPLICANT: Kongsbak, Lars
TITLE OF INVENTION: No. 6187580el Pectate Lyases
FILE REFERENCE: 5378-200-US
CURRENT APPLICATION NUMBER: US/09/198, 955A
CURRENT FILING DATE: 1998-11-24
PRIOR APPLICATION NUMBER: 1343/97
PRIOR FILING DATE: 1997-11-24
PRIOR APPLICATION NUMBER: 1344/97
PRIOR FILING DATE: 1997-11-24
PRIOR APPLICATION NUMBER: 60/067, 249
PRIOR FILING DATE: 1997-12-02
PRIOR APPLICATION NUMBER: 60/067, 240
PRIOR FILING DATE: 1997-12-02
PRIOR APPLICATION NUMBER: 09/073, 684
PRIOR FILING DATE: 1998-05-06
PRIOR APPLICATION NUMBER: 09/184, 217
PRIOR FILING DATE: 1998-11-02
NUMBER OF SEQ ID NOS: 32
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 12
LENGTH: 493
TYPE: PRT
ORGANISM: Clostridium thermocellum
US-09-198-955A-12

Query Match 10.1%; Score 415.5; DB 4; Length 493;
Best Local Similarity 27.3%; Pred. No. 7.3e-21;
Matches 112; Conservative 71; Mismatches 126; Indels 101; Gaps 14;

QY 247 WVDQMSAYVKSIDPNHVMVSGDEGYIG-----STQSGWPY-----NDPS 287
DB 135 WVDHNELY-----HSLNV-DKDYDGLFDVKKRDAEYITFSWNYVHDGKMSMLMGSSDS 186
QY 288 DGVN-----NNALLRVKNIDFGTYHLYPNYWQONADGQWIKDHIANAAIGKP 337
DB 187 DYNRTITTFHHNFENLSRVSPFRGEGHIYNNYFNKIIDSG-----229
QY 338 TILFEFCWOTPDSDSVYQVOTQTQVTRTNGEAGWFMMLAGN--VNGQ---PYPNYDGENVY 392
DB 230 -INSRMCARIRIENLFPENAKDPIVSWSSPGYVHWNNKFNVSRCGSMPTTSTTTNNPP 288
QY 393 YPSTATV-----LASEALAISTGTSTPPSPSSPSSPSSPSSPSSPSSPSSPSSPSS 447
DB 289 YSYSLDNVDNVKSVIKQKONAGVGKIQRRPPTP--TPTSPPS-----326
QY 448 PSPSSSPVSGVGVQYKKNDSAPCDNOIKPGLQLVNTGSSVDLSTVTYVWFTTRDGGSS 507
DB 327 ---ANTPVSNLKVFEFYNNSPDTTNSINPQFKVTNTGSSAIDLSKLTLYRYTYVDGQKD 383
QY 508 TLVYNCDAAM-----CGGNIRASFGSVNPATPTADTYLQLSFTGTLAAGSGTGE 558
DB 384 QTFW-CDHAAIIGSNGSYNGITSNVKCTFVKMSSTNNADTYLEISFTGTLPEGAHV-Q 441
QY 559 IQNRVKNKSDNSFDETNDYSGTNTAFQDWTKVTYVYNGRLVWGTEPSGT 608
DB 442 IQGRFAKNDSNTQSDNYSFKRSQFVENDQVTAYLNGVLVWGKEPGGS 491

RESULT 12
US-09-694-531-12
Sequence 12, Application US/09694531
Patent No. 6368843
GENERAL INFORMATION:
APPLICANT: Andersen, Lene N.
APPLICANT: Schuelein, Martin
APPLICANT: Lange, Niels E.
APPLICANT: Bjornvad, Mads E.
APPLICANT: Molier, Soren
APPLICANT: Glad, Sanne O. S.
APPLICANT: Kauppinen, Markus S.
APPLICANT: Schnorr, Kirk
APPLICANT: Kongsbak, Lars

TITLE OF INVENTION: No. 6368843el Pectate Lyases
FILE REFERENCE: 5378-200-US
CURRENT APPLICATION NUMBER: US/09/694, 531
CURRENT FILING DATE: 2000-10-23
PRIOR APPLICATION NUMBER: 09/198, 955
PRIOR FILING DATE: 1998-11-24
PRIOR APPLICATION NUMBER: 1343/97
PRIOR FILING DATE: 1997-11-24
PRIOR APPLICATION NUMBER: 1344/97
PRIOR FILING DATE: 1997-11-24
PRIOR APPLICATION NUMBER: 60/067, 249
PRIOR FILING DATE: 1997-12-02
PRIOR APPLICATION NUMBER: 60/067, 240
PRIOR FILING DATE: 1997-12-02
PRIOR APPLICATION NUMBER: 09/073, 684
PRIOR FILING DATE: 1998-05-06
PRIOR APPLICATION NUMBER: 09/184, 217
PRIOR FILING DATE: 1998-11-02
NUMBER OF SEQ ID NOS: 32
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 12
LENGTH: 493
TYPE: PRT
ORGANISM: Clostridium thermocellum
US-09-694-531-12

Query Match 10.1%; Score 415.5; DB 4; Length 493;
Best Local Similarity 27.3%; Pred. No. 7.3e-21;
Matches 112; Conservative 71; Mismatches 126; Indels 101; Gaps 14;

QY 247 WVDQMSAYVKSIDPNHVMVSGDEGYIG-----STQSGWPY-----NDPS 287
DB 135 WVDHNELY-----HSLNV-DKDYDGLFDVKKRDAEYITFSWNYVHDGKMSMLMGSSDS 186
QY 288 DGVN-----NNALLRVKNIDFGTYHLYPNYWQONADGQWIKDHIANAAIGKP 337
DB 187 DYNRTITTFHHNFENLSRVSPFRGEGHIYNNYFNKIIDSG-----229
QY 338 TILFEFCWOTPDSDSVYQVOTQTQVTRTNGEAGWFMMLAGN--VNGQ---PYPNYDGENVY 392
DB 230 -INSRMCARIRIENLFPENAKDPIVSWSSPGYVHWNNKFNVSRCGSMPTTSTTTNNPP 288
QY 393 YPSTATV-----LASEALAISTGTSTPPSPSSPSSPSSPSSPSSPSSPSSPSSPSS 447
DB 289 YSYSLDNVDNVKSVIKQKONAGVGKIQRRPPTP--TPTSPPS-----326
QY 448 PSPSSSPVSGVGVQYKKNDSAPCDNOIKPGLQLVNTGSSVDLSTVTYVWFTTRDGGSS 507
DB 327 ---ANTPVSNLKVFEFYNNSPDTTNSINPQFKVTNTGSSAIDLSKLTLYRYTYVDGQKD 383
QY 508 TLVYNCDAAM-----CGGNIRASFGSVNPATPTADTYLQLSFTGTLAAGSGTGE 558
DB 384 QTFW-CDHAAIIGSNGSYNGITSNVKCTFVKMSSTNNADTYLEISFTGTLPEGAHV-Q 441
QY 559 IQNRVKNKSDNSFDETNDYSGTNTAFQDWTKVTYVYNGRLVWGTEPSGT 608
DB 442 IQGRFAKNDSNTQSDNYSFKRSQFVENDQVTAYLNGVLVWGKEPGGS 491

RESULT 13
US-09-670-141-10
Sequence 10, Application US/09670141
Patent No. 6429000
GENERAL INFORMATION:
APPLICANT: Andersen, Lene N.
APPLICANT: Schuelein, Martin
APPLICANT: Lange, Niels Erik K.
APPLICANT: Bjornvad, Mads E.
APPLICANT: Schnorr, Kirk
TITLE OF INVENTION: Pectin Degrading Enzymes From Bacillus
TITLE OF INVENTION: Licheniformis
FILE REFERENCE: 5377-200-US
CURRENT APPLICATION NUMBER: US/09/670,141

```

RESULT 14
PCT-US95-10813-9
; Sequence 9, Application PC/TUS9513813
; GENERAL INFORMATION:
; APPLICANT: Yeda Research and Development Co. Ltd.
; APPLICANT: Ramot University Authority for Applied
; APPLICANT: Research and Industrial Development Ltd..
; APPLICANT: Technion Research and Development Foundation Ltd.
; APPLICANT: Bayer, Edward A.
; APPLICANT: Morag, Ely
; APPLICANT: Wilchek, Meir
; APPLICANT: Lamed, Raphael
; APPLICANT: Shoham, Yuval
; TITLE OF INVENTION: MODIFIED CELLULOSE-BINDING DOMAIN (CBD)
; TITLE OF INVENTION: PROTEINS AND USE THEREOF
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESS: Brodsky and Nelmark
; STREET: 419 Seventh Street N.W., Ste. 300
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20004
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk

```


According to the Pre Publication Rules, every patent application received by the United States Patent and Trademark Office after November 29, 2000 will be pre-published at eighteen months from the effective filing date. When the application is published the contents, including the sequences, will become prior art.

Two new databases have been created to hold the pre-published sequences:

Published_Applications_NA contains nucleic acid sequences; the search results will have the extension **.rnpb**.

Published_Applications_AA contains amino acid sequences; the search results will have the extension **.rapb**.

Each pre-published application is given a unique Publication Number. An example of a Publication Number is US20021234567A1. The "US" indicates the application was a U.S. application. The first 4 digits show the calendar year the application was published. The next 7 digits represent when the application was published. This 7-digit number starts at zero at the beginning of each calendar year. Each application published is given the next number in order. The "A" indicates a utility patent application and the "1" shows that this was the first time the application had been published. If the applicants submit changes to the application, they may request that the changed application be published again. In such instances, the "1" at the end of the number would be replaced by a "2".

Sequences in the PGPub database are public information; it is permissible to leave these results in the case.

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 13, 2002, 11:55:11 : Search time 14.7802 Seconds
(without alignments)
776.473 Million cell updates/sec

Title: US-09-917-378-1

Perfect score: 4122

Sequence: 1 MGLVRRPARAFVATAAGTAV.....GFNGSYSGTNTAPTTLTCTAS 762

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 97044 seqs, 15060890 residues

Total number of hits satisfying chosen parameters: 97044

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published_Applications_AA:*

1: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB pep.*
2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB pep.*
3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB pep.*
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB pep.*
5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB pep.*
6: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB pep.*
7: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB pep.*
8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB pep.*
9: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB pep.*
10: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB pep.*
11: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB pep.*
12: /cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB pep.*
13: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB pep.*
14: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	813	19.7	680	9	US-10-121-032-25
2	812	19.7	663	9	US-10-121-032-61
3	696.5	16.9	562	10	US-09-981-900B-5
4	494	12.0	427	9	US-09-850-982B-2
5	415.5	10.1	493	12	US-10-072-152-12
6	245	5.9	1367	10	US-09-801-368-108
7	238	5.8	2344	10	US-09-815-242-12713
8	218.5	5.3	5179	10	US-09-922-217-1068
9	218.5	5.3	5179	10	US-09-833-263-1068
10	199.5	4.8	371	10	US-09-739-861A-1
11	199.5	4.8	371	10	US-09-795-583-1
12	198.5	4.8	386	10	US-09-739-861A-5
13	198.5	4.8	386	10	US-09-795-583-5
14	189	4.6	596	12	US-10-006-867-100
15	189	4.6	596	12	US-10-052-586-310
16	186	4.5	560	10	US-09-815-242-13057
17	183.5	4.5	502	10	US-09-815-242-5904
18	182.5	4.4	957	10	US-09-922-217-1085
19	182.5	4.4	957	10	US-09-833-263-1065

Sequence 10936, A
Sequence 36047, A
Sequence 39507, A
Sequence 428, App
Sequence 2, Appl
Sequence 10932, A
Sequence 394, App
Sequence 395, App
Sequence 2, Appl
Sequence 2, Appl
Sequence 176, App
Sequence 12, Appl
Sequence 2, Appl
Sequence 104, App
Sequence 52, Appl
Sequence 2, Appl
Sequence 62, Appl
Sequence 32, Appl
Sequence 2, Appl
Sequence 25, Appl
Sequence 95, Appl
Sequence 1, Appl
Sequence 154, App
Sequence 154, App

175 4.2 1333 10 US-09-815-242-10936
21 172.5 4.2 688 10 US-09-864-761-36047
22 164 4.0 113 10 US-09-864-761-39507
23 163.5 4.0 503 10 US-09-801-368-424
24 163.5 4.0 605 10 US-09-801-368-428
25 162 3.9 881 10 US-09-816-860A-2
26 161.5 3.9 802 10 US-09-823-240-2
27 159.5 3.9 1031 10 US-09-815-242-10932
28 154 3.7 1723 10 US-09-841-132-394
29 154 3.7 1723 10 US-09-841-132-395
30 152.5 3.7 945 8 US-08-965-272-2
31 152.5 3.7 945 12 US-10-121-882-2
32 149.5 3.6 832 10 US-09-729-674-176
33 148.5 3.6 659 12 US-10-090-624-12
34 148 3.6 572 10 US-09-943-075A-2
35 147 3.6 1140 12 US-10-124-557-104
36 147 3.6 1363 12 US-10-124-557-52
37 147 3.6 1404 12 US-10-124-557-2
38 147 3.6 1404 12 US-10-124-557-62
39 146.5 3.6 2843 8 US-08-681-219-32
40 145 3.5 553 10 US-09-888-224-2
41 144.5 3.5 1744 9 US-10-108-605-25
42 144 3.5 1016 12 US-10-007-693-95
43 143.5 3.5 467 10 US-09-863-547B-1
44 143.5 3.5 693 10 US-09-752-639-154
45 143.5 3.5 693 10 US-09-984-198-154

ALIGNMENTS

RESULT 1

Sequence 25, Application US/10121032

Patent No. US2002015550A1

GENERAL INFORMATION:

APPLICANT: Bylina, Edward J.

TITLE OF INVENTION: GLYCOSIDASE ENZYMES

NUMBER OF SEQUENCES: 72

CORRESPONDENCE ADDRESS:

ADDRESSEE: Gray Cary Ware & Freidenrich LLP

STREET: 4365 Executive Drive, Suite 1600

CITY: San Diego

STATE: CA

COUNTRY: USA

ZIP: 92121

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette

OPERATING SYSTEM: Windows95

SOFTWARE: FastSeq for Windows Version 2.0

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/10121,032

FILING DATE: 09-Apr-2002

CLASSIFICATION: <unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US/09/134,078

FILING DATE: 13-AUG-1998

APPLICATION NUMBER: 08/949,026

FILING DATE: 10-OCT-1997

APPLICATION NUMBER: 60/056,916

FILING DATE: 06-DEC-1996

ATTORNEY/AGENT INFORMATION:

NAME: Halle, Lisa A.

REGISTRATION NUMBER: 38,347

REFERENCE/DOCKET NUMBER: 09010/024002

TELEPHONE: 858/677-1456

TELEFAX: 658/677-1465

INFORMATION FOR SEQ ID NO: 25:

SEQUENCE CHARACTERISTICS:

LENGTH: 680 amino acids

TYPE: amino acid

;
; MOLECULE TYPE: linear
; FRAGMENT TYPE: Internal
; SEQUENCE DESCRIPTION: SEQ ID NO: 25:
US-10-121-032-25

Query Match 19.7%; Score 813; DB 9; Length 680;
Best Local Similarity 41.5%; Pred. No. 2.le-37;
Matches 173; Conservative 65; Mismatches 135; Indels 44; Gaps 13;

Oy 19 AVAAATLGSITMPSATAAPAGVTASGGQFVLNGLPYRYGTTNNYLSYOSHADVDDVL 78
: :
Db 12 SVSAEFLLLIVELSPLEASDEFKVKENGKALNGKEFRFIGSNYYMHYKSNMGIDSVL 71

Oy 79 AKQAAMNLSVIRTWGCFIDIGSLDGSVPTIDGNKNGFYFOYWDPTGAPAYNDG----PTG 134
: :
Db 72 ESADMGKIVLRIMWGF-----LDGESYCRDKN-----TYMHPEPGVGVPEGISNAQSG 120

Oy 135 LQGLDYAIASAAHGLRIVVLTNDWKEFGGMDQYDKWGLPYHDNFYDPTQOAYKNW 194
: :
Db 121 FERLDYTVAKAKELGIKLVIVLNWDDFGCMNQYVRWFGGTHHDDFYRDEKIKEEYKY 180

Oy 195 VNHLNRVNSITGVYKNDPTIFANELANPRCVSGSLPTSGTCTQATIVNWDOMSA 254
: :
Db 181 VSFLVNHVNTYGVYREPTIMAWELANPRC-----ETDKSGN-----TLVEWVKEMSSY 232

Oy 255 VKSIDPNHMSVGDGEF---YIGSTQSG---WPYNDPSDGVNALLRVKNIDFCTYHL 308
: :
Db 233 IKSLDPNHLVAVGDGEFFSNTEGFKPYGGEAEWAYNGWS-GVDWKKLLSIETVDFGTGPHL 291

Oy 309 YPNYWGQN----ADMGTQWIKDHIAANAAGIKPTILEEFG--WOTP-DRDSVYQVOTV 361
: :
Db 292 YPSHMGVSPENYAQWAKWIEDHIAKEIGKPVVLEEYGIKPSAPVNRRTAIYRLNNDLV 351

Oy 362 RTNCEAGNFWMLAGNVGNOP-----YPNVDFGNVYPSSTATVLAASE-ALAI 411
: :
Db 352 YDLGGGAMFWMLAGIGSDRDERGYYPDYDGFRIVNDSDPEALIREYAKLFNTG 408

RESULT 2
US-10-121-032-61
; Sequence 61, Application US/10121032
; Patent No. US20020155550A1
; GENERAL INFORMATION:
; APPLICANT: Bylina, Edward J.
; TITLE OF INVENTION: GLYCOSIDASE ENZYMES
; NUMBER OF SEQUENCES: 72
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Gray Cary Ware & Freidenrich LLP
; STREET: 4365 Executive Drive, Suite 1600
; CITY: San Diego
; STATE: CA
; COUNTRY: USA
; ZIP: 92121
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: Windows95
; SOFTWARE: FastSEO for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/121.032
; FILING DATE: 09-Apr-2002
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/09/134.078
; FILING DATE: 13-AUG-1998
; APPLICATION NUMBER: 08/949,026
; FILING DATE: 10-OCT-1997
; APPLICATION NUMBER: 60/056,916
; FILING DATE: 06-DEC-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Hallie, Lisa A.
; REGISTRATION NUMBER: 38,347

;
; REFERENCE/DOCKET NUMBER: 09010/024002
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 858/677-1456
; TELEFAX: 858/677-1465
; INFORMATION FOR SEQ ID NO: 61:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 663 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; FRAGMENT TYPE: Internal
; SEQUENCE DESCRIPTION: SEQ ID NO: 61:
US-10-121-032-61

Query Match 19.7%; Score 812; DB 9; Length 663;
Best Local Similarity 42.8%; Pred. No. 2.3e-37;
Matches 169; Conservative 61; Mismatches 121; Indels 44; Gaps 13;

Oy 41 FVTASGGQFVLNGLPYRYGTTNNYLSYOSHADVDDVLAKQAAMNLSVIRTWGCFIDIGSL 100
: :
Db 17 FVKVENGKALNGKEFRFIGSNYYMHYKSNMGIDSVLESARDMGIKVLRIMWGF-----L 71

Oy 101 DGSVPTIDGNKNGFYFOYWDPTGAPAYNDG---PTGLQGLDYAIASAAHGLRIVVVL 156
: :
Db 72 DGESYCRDKN-----TYMHPEPGVGVPEGISNAQSGFERLDYTVAKAKELGIKLVIVL 125

Oy 157 TNDWKEFGGMDQYDKWGLPYHDNFYDPTQOAYKNVNHLLNRVNSITGVYTKNDPTI 216
: :
Db 126 VNNWDDFGCMNQYVRWFGGTHHDDFYRDEKIKEEYKYVSLVNHVNTYGVYREPTI 185

Oy 217 FAWELANPRCVSGSLPTSGTCTQATIVNWDOMSAVYKSIDPNHMSVGDGEF---YI 273
: :
Db 186 MAWELANPRC-----ETDKSGN-----TLVEWVKEMSSYIKSLDPNHLVAVGDGEFFSNTE 237

Oy 274 GSTQSG---WPYNDPSDGVNALLRVKNIDFGTYHLYPNYWGQN---ADMGTQWIKD 326
: :
Db 238 GFKPYGGEAEWAYNGWS-GVDWKKLLSIETVDFGTGTHLYPSHMGVSPENYAQWAKWIED 296

Oy 327 HIANAAGIKPTILEEFG--WOTP-DRDSVYQVOTVOTVNTNCEAGNFWMLAGNVGNOP- 382
: :
Db 297 HIKIAKEIGKPVVLEEYGIKPSAPVNRRTAIYRLNNDLVYDLGGGAMFWMLAGIGESD 356

Oy 383 -----YPNVDFGNVYPSSTATVLAASE-ALAI 411
: :
Db 357 DERGYYPDYDGFRIVNDSDPEALIREYAKLFNTG 391

RESULT 3
US-09-981-900B-5
; Sequence 5, Application US/09981900B
; Patent No. US20020138878A1
; GENERAL INFORMATION:
; APPLICANT: Sticklen, Masomeh B
; APPLICANT: Maqbool, Shahina B
; APPLICANT: Dale, Bruce E
; TITLE OF INVENTION: TRANSGENIC PLANTS CONTAINING LIGNINASE AND CELLULOSE WHICH DEGRAD CELLULOSE TO FERMENTABLE SUGARS
; FILE REFERENCE: MSU 4.1-539
; CURRENT APPLICATION NUMBER: US/09/981.900B
; CURRENT FILING DATE: 2002-03-18
; PRIOR APPLICATION NUMBER: 60/242,408
; PRIOR FILING DATE: 2000-10-20
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 5
; LENGTH: 562
; TYPE: PRT
; ORGANISM: Acidothermus cellulolyticus
US-09-981-900B-5

Query Match 16.9%; Score 696.5; DB 10; Length 562;
Best Local Similarity 27.3%; Pred. No. 3.6e-31;
Matches 221; Conservative 93; Mismatches 194; Indels 301; Gaps 34;

QY 4 VRRPARAEVATAAGTAAVA---AATLGSITMP-SATAAPAGFVTASGQFV-LNGLPYRY 58
Db 5 LRRVPCSRVHLRVGVVAVLALVAALANLAVPRPARAAGGVTGHTSGREILDANNVPVRI 64
QY 59 GCTN-----NYLSYQSHADVDVLAQAQANLNVIRTWGFDIGSLDGSVPTIDGNK 111
Db 65 AGINWPFETCNYVHCLWSRYSRLDQIKSLGYNTIRL-PYSDDILKPGTMP-----NS 119
QY 112 NGFYQYWDPSGAPAYNDGPTGLQGLDYAIAASAAHGLRVIVLTDNWKKEGGMQDYOK 171
Db 120 INFOQNOD-----LOGILSLOWMDKIVAYAGIGIRII--LDRHRPDCSG--QSAL 167
QY 172 WYGLPYHDFYDPTQQAQYKNNVHLLNRVNSITGVTKNDPTIFAWELANEPR---CV 228
Db 168 WY-----TSSVSEATWISDL-----QALAQRYKGNPTVVGFDLHNEPHDPACW 210
QY 229 GSGTLPTSTCTQATIVNN---VDQMSAVYKSIDPNHMYSV-----GDEGFY IGSTG 278
Db 211 GCGD-PS-----IDWRLAERAGNAVLNPNLLIFVEGVQSYNGDSYVMWGNLQ 260
QY 279 SG-WPYNDPSDGVNALLRVKN-----IDFCTYHLYPNYQONAD-----WGTOW 323
Db 261 AGOYP-----VVLNPNRLVYSAHDYAT-SVYPOTWFSDBTFFNNPFGIWNKNW 308
QY 324 ---IKDHIAANAAGIKPTILEEFGMOTPDORSVYQWTOTVIRTNNGEAGNFWMLACNVNG 380
Db 309 GYLFNQIA-----PWLGEFG---TTLQSTTDQTLKTL-----340
QY 381 QYPYNDGVNYYPSSTATVLAASEALAISTGTPPPSPSSPSSPSSPSSPSSPSSPSSP 440
Db 341 -----340
QY 441 SASSPSPSSPSSPVGGVKVOYKKNDSAPGDNOIKPGLQVNTGSSVDLSTVTYRYWF 500
Db 341 -----VQY-----LRP-----TAQY---350
QY 501 TRDGSSTLVNCDMAAMCGNIRASFGSVNPAATPTADTYLQLSFTGTLAAGGSTGEIQ 560
Db 351 -----CADSFQMTFWSNPDS-----GDTGGIL 373
QY 561 NRVNKSQNSNFDETNDYSGTNTATQDWTKVTVYVNGRLVWGTPESGTSPSPSPSP 620
Db 374 ----KDDMOTVDTVKD-----GYLAPIKSSIF-----DPYGASASPSOSPSPSV 413
QY 621 SP 673
Db 414 SP 473
QY 674 DWGSGFTATVTVNTGSRATSGMTVAMSGFGNQTIVYNTALTQSGASVATNLSYNNV 733
Db 474 DWNGFTVTVAVTNSGVSATKTWTVSWTEGNGQITNSNNAATQNGQSVTARNKSYNNV 533
QY 734 IOPGOSTTFGNGSVSGTNTATPLTCTAS 762
Db 534 IOPGQNTTFGFOASVTGTSNAAPTACAAS 562

RESULT 4
US-09-850-982B-2
; Sequence 2, Application US/09850982B
; Patent No. US20020166145A1
; GENERAL INFORMATION:
; APPLICANT: Nestec S.A.
; TITLE OF INVENTION: COFFEE MANNANASE
; FILE REFERENCE: 88265-4025
; CURRENT APPLICATION NUMBER: US/09/850.982B
; CURRENT FILING DATE: 2001-05-08
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 2
; LENGTH: 427
; TYPE: PRT

; ORGANISM: Coffea arabica
US-09-850-982B-2
Query Match 12.0%; Score 494; DB 9; Length 427;
Best Local Similarity 31.6%; Pred. No. 2.6e-20;
Matches 125; Conservative 66; Mismatches 139; Indels 66; Gaps 16;
QY 35 TAAPAGFVTASGQFVNLGLPYRYGCTNNYYLSY-----QSHADVDVLAQAQANLNSVI 89
Db 32 SSSASRFQTRGTRFVLGGYFFFFNGFNSYMMHVAAPSERIKISNVFREAAATGLTVC 91
QY 90 RTWGFIDIGSLDGSVPTIDGNKNGFYQYWDPSGAPAYNDGPTGLQGLDYAIAASAAHG 149
Db 92 RTWAFSDGG--DRALQMSPG-----VYDE--RVFOALDFVYSEARKYC 130
QY 150 LRVIVVLTNDWKEFGMDQYDKW---YGLPYH--DNFYDPTQQAQYKNNVHLLNRVNS 204
Db 131 VHLILSLTNNYKDFGRTQYVTVMAKNAGVQVNSDDDFYTKNAVGYKNNHKKVLTRINT 190
QY 205 ITGVTVYKNDPTIFAWELANEPRCVSGSLTPTSGTCTQATIVNVDQMSAVYKSIDPNHMY 264
Db 191 ISRVATKDDPTVMAMELINEPRC-----QVDFSG-----KTLNAWVQEMATVYKSLDNKHLL 242
QY 265 SVGDEGFYIGSTQGS---GWPNYNDPSGDVNNALLRVKNIDFGTYHLYPNYMGONADWGT 321
Db 243 EIGMEGFYIGDSMPGKKOYNPGYQVGTDFITNNL---IKEIDFATIHAYPDIWLSGSDGA 299
QY 322 Q-----WTKDHIAANAAGI-KKPTILEEFGMOTPD-----RDS-----VYQWTOTVIRT 363
Db 300 QNMFMRRMTSHSTSKTILKKPLVLAETCKSKKOPGYSLYARESPMAAIYGDYIRFARR 359
QY 364 NCEAGNFWMLACNVNCPYPNYDGFNVYPSSTAT 399
Db 360 GGIAGLVWOLAE-QMOPYA--DGYEIVLSQNPST 392

RESULT 5
US-10-072-152-12
; Sequence 12, Application US/10072152
; Patent No. US20020142438A1
; GENERAL INFORMATION:
; APPLICANT: Andersen, Lene N.
; APPLICANT: Schulein, Martin
; APPLICANT: Lange, Niels E.
; APPLICANT: Bjornvad, Mads E.
; APPLICANT: Moller, Soren
; APPLICANT: Glad, Sanne O. S.
; APPLICANT: Kauppinen, Markus S.
; APPLICANT: Schnorr, Kirk
; APPLICANT: Kongsbak, Lars
; TITLE OF INVENTION: NO. US20020142438A1el Peclate Lyases
; FILE REFERENCE: 5378.200-US
; CURRENT APPLICATION NUMBER: US/10/072,152
; CURRENT FILING DATE: 2002-02-07
; PRIOR APPLICATION NUMBER: US/09/198,955
; PRIOR FILING DATE: 1998-11-24
; PRIOR APPLICATION NUMBER: 1343/97
; PRIOR FILING DATE: 1997-11-24
; PRIOR APPLICATION NUMBER: 1344/97
; PRIOR FILING DATE: 1997-11-24
; PRIOR APPLICATION NUMBER: 60/067,249
; PRIOR FILING DATE: 1997-12-02
; PRIOR APPLICATION NUMBER: 60/067,240
; PRIOR FILING DATE: 1997-12-02
; PRIOR APPLICATION NUMBER: 09/073,684
; PRIOR FILING DATE: 1998-05-06
; PRIOR APPLICATION NUMBER: 09/184,217
; PRIOR FILING DATE: 1998-11-02
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12
; LENGTH: 493
; TYPE: PRT

Query Match 5.9%; Score 245; DB 10; Length 1367;
Best Local Similarity 26.6%; Pred. No. 2.8e-06;
Matches 113; Conservative 56; Mismatches 172; Indels 84; Gaps 14;

```

US-09-815-242-12713
: Sequence 12713, Application US/09815242
: Patent No. US20020061569A1
: GENERAL INFORMATION:
: APPLICANT: Haselbeck, Robert
: APPLICANT: Ohlsen, Karl L.
: APPLICANT: Zyskind, Judith W.
: APPLICANT: Wall, Daniel
: APPLICANT: Trawick, John D.
: APPLICANT: Carr, Grant J.
: APPLICANT: Yamamoto, Robert T.
: APPLICANT: Xu, H. Howard
: TITLE OF INVENTION: Identification of Ess
: TITLE OF INVENTION: Prokaryotes
: FILE REFERENCE: ELITRA-011A
: CURRENT APPLICATION NUMBER: US/09/815,242
: CURRENT FILING DATE: 2001-03-21
: PRIOR APPLICATION NUMBER: 60/191,078
: PRIOR FILING DATE: 2000-03-21
: PRIOR APPLICATION NUMBER: 60/206,848
: PRIOR FILING DATE: 2000-03-23
: PRIOR APPLICATION NUMBER: 60/207,727
: PRIOR FILING DATE: 2000-05-26
: PRIOR APPLICATION NUMBER: 60/242,578
: PRIOR FILING DATE: 2000-10-23
: PRIOR APPLICATION NUMBER: 60/253,625
: PRIOR FILING DATE: 2000-11-27
: PRIOR APPLICATION NUMBER: 60/257,931
: PRIOR FILING DATE: 2000-12-22
: PRIOR APPLICATION NUMBER: 60/269,308
: PRIOR FILING DATE: 2001-02-16
: NUMBER OF SEQ ID NOS: 14110
: SOFTWARE: FastSeq for Windows Version 4.0
: SEQ ID NO 12713
: LENGTH: 2344
: TYPE: PRT
: ORGANISM: Staphylococcus aureus
US-09-815-242-12713

```

[illegible]

RESULT 8
US-09-922-217-1068
; Sequence 1068, Application US/09922217

```

: Patent No. US20020076414A1
:
: GENERAL INFORMATION:
:
: APPLICANT: Xu, Jtjangchun
: APPLICANT: Lodes, Michael
: APPLICANT: Secretist, Heat
: APPLICANT: Benson, Darin
: APPLICANT: Meagher, Madeir
: APPLICANT: Stolk, John A
: APPLICANT: Wang, Tongtong
: APPLICANT: Jiang, Yuqiu
: APPLICANT: Smith, Carolee
: APPLICANT: King, Gordon
: APPLICANT: Wang, Aijun
: APPLICANT: Clapper, Jona
:
: TITLE OF INVENTION: COMPO
:
: FILE REFERENCE: 210121.47
:
: CURRENT APPLICATION NUMBER:
:
: CURRENT FILING DATE: 200
:
: NUMBER OF SEQ ID NOS: 112
:
: SOFTWARE: Fast-SEQ for Win
:
: SEQ ID NO 1068
:
: LENGTH: 5179
:
: TYPE: PRT
:
: ORGANISM: Homo sapiens
:
: US-09-922-217-1068

```

Query Match	5.3%	Score 218.5	DB 10	Length 5179
Best Local Similarity	22.6%	Pred. No. 0.00029		
Matches 98	Conservative	69	Mismatches 159	Indels 107
				Gaps 19

Qy	394	PSSTATVLA	SEALAI	STGTSPP	--SPSSPSSSSP	-----SPSPASPSA	-----S	439
Db	1657	PSPTTTP	TPSP--PI	TTTTTTP	PTTPSP	PIITTPSP	TTTTTTP	1714
Qy	440	PSASSSP	SPSPSS--	-----SPVSG	GKVOYKND	SAP--GDNQ	IKPGL	488
Db	1715	PSSTT	TPSPPTT	PTTTPSP	PTTTP	TTTTP	SSPLTT	1774
Qy	489	VDLST	-----VTVRYW	-----FTR	DGSSLS	LVYN-C--DWA	AGCGN	532
Db	1775	TTTPT	PCVPLC	NWTKL	DGSKPN	FKPGD	TELIG	1829
Qy	533	ATPTAD	YIQL	SFTG	TLAGG	STGEI	QNRV	587
Db	1830	DVP	-----IGOL	GTVC	DVSVGL	ICKN	EDOK	1882
Qy	588	WTKVTV	VNGLV	WGTPE	SGTSP	STPSP	PTPSP	636
Db	1883	TTMTT	TTT	TENP	TPPTT	PTT	TTTTV	1942
Qy	637	SPSP	-----SPSP	TPSP	SPSP	SPSP	SSVSG	689
Db	1943	TTPT	TTTTV	TP	TP	TP	TG	1984
Qy	690	SRATSG	MTVA	MSFG	NO	TVN	WNTAL	749
Db	1985	TOPT	TTTT	PT	-----TTTT	TTTT	PT	2033
Qy	750	GTNTAP	L	CTAS	762			
Db	2034	PTTPT	TTTT	TT	2046			

RESULT 9
US-09-833-263-1068
: Sequence 1068, Application US/09833263
: Patent No. US20030110547A1
: GENERAL INFORMATION:
: APPLICANT: Wang, Aijun
: APPLICANT: Clapper, Jonathan D.
: APPLICANT: Stolk, John A.
: APPLICANT: Meagher, Madeleine J.

```
; TITLE OF INVENTION: COMPOUNDS FOR IMMUNOTHERAPY AND
; FILE REFERENCE: 210121.471C12
; CURRENT APPLICATION NUMBER: US/09/833.263
; CURRENT FILING DATE: 2001-04-10
; NUMBER OF SEQ ID NOS: 1093
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 1068
; LENGTH: 5179
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-09-833-263-1066

Query Match      5.3%   Score 218.5; DB 10; Length 5179;
Best Local Similarity 22.6%; Pred. No. 0.00029;
Matches 98; Conservative 69; Mismatches 159; Indels 107; Gaps 19;

Qy 394 PSSTATVLASALAICTGCTSPSP--SPSSPSSSPSP-----SPSPSPSPSA-----S 439
   || | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1657 PSPTTTTTPSP--PITTTTTPPTTTPSSPITTTTTPSPPTTMTTTPSPPTTTPSSPITTTT 1714
   || | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 440 PSASSSPSPSPSS-----SPVSGVKVQYKNDSP-GDNQIKPGLQLVNTGSSS 488
   || | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1715 PSSTTTSPPTTTPSPPTTTPSPPTTMTTLPPTTSSPLTTTLPSPPTTTPSPPTTTPSP 1774
   || | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 489 VDLST---VTVRYW-----FTRDGSSTLVYN-C--DMAAMCGGNIRASFGSYNP 532
   || | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1775 TTPTTTPCPLCNMTGLDCKFNFKPGDTELIGDVGCGMAA----NISCR-ATMYP 1829
   || | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 533 ATPTADTYLQLSFTGGTLAAGSGTGEIQNRVKNKSDMSNFDNDSYGTNT-----AFQD 587
   || | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1830 DVP-----IGQJGQIVVCVDSVGLICKNEDQKPGGIVPMAFCLNYEINVCCECVTP 1882
   || | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 588 WTKVTYVYNGRLVNGTPEPSTPSPSTPSPSP-----SPTPSPSSSPSP 636
   || | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1883 TTMTTTTENTPPTTPITTTTTPPTPTGTPTTPTTITTTTTPPTPTPTGTGTP 1942
   || | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 637 SPSP-----SPSTPSPSPSPSPSPSSVSSGVCGRATYVYVNSDMGSGFTATVTVNTG 689
   || | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1943 TTTPTTTTTPPTPTGTPTTPTTITTT-----TVTPTPTPTG 1984
   || | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 690 SRATSGMTVAWSFGNGTNTVYWNALTQSGASVTATNLSYNNVLPORCOSTTFGFNGSYS 749
   || | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1985 TOTPTTPTIT-----TTTTVTPTPTGTGTPTTPTTPTTPTTPTPTPTPTPTPTPTPT 2033
   || | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 750 GTNTAPTITCTAS 762
   || | | | |
Db 2034 PTTTPTTITTTVT 2046
   || | | | |

RESULT 10
US-09-739-861A-1
; Sequence 1, Application US/09739861A
; Patent No. US2002007692A1
; GENERAL INFORMATION:
; APPLICANT: Jones, Brien E.
; APPLICANT: van der Kleij, Wilhelmus A.H.
; APPLICANT: van Solingen, Piet
; APPLICANT: Weyler, Walter
; APPLICANT: Goedegebuur, Frits
; TITLE OF INVENTION: No. US2002007692A1 Cellulase Producing Actinomycetes,
; FILE REFERENCE: GC540-3
; CURRENT APPLICATION NUMBER: US/09/739, 861A
; CURRENT FILING DATE: 2000-12-18
; PRIOR APPLICATION NUMBER: US 09/321,981
; PRIOR FILING DATE: 1999-05-28
; PRIOR APPLICATION NUMBER: US 09/104,308
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: US 08/974,042
; PRIOR FILING DATE: 1997-11-19
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: FastSeq for Windows Version 3.0
```

```
; SEQ ID NO 1
; LENGTH: 371
; TYPE: PRT
; ORGANISM: Unknown
; FEATURE:
; OTHER INFORMATION: Nearest "neighbor" - Streptomycetes
; OTHER INFORMATION: thermovibaceus
US-09-739-861A-1

Query Match      4.8%   Score 199.5; DB 10; Length 371;
Best Local Similarity 20.8%; Pred. No. 0.00021;
Matches 101; Conservative 49; Mismatches 153; Indels 183; Gaps 22;

Qy 303 FGTVHLVYNGQONADWCTQWIKOHIAAAIGKPTILEEFCWQTPDRDSYVQTWTQTVR 362
   || | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 39 YGTTTIQIRYVQNNRWCT-----SATQCIN 64
   || | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 363 TNGEAGNFMWLAGNV--NGQP--YPN-YDGFNVYPSSTATVLAASEALAICTGTPSPPS 417
   || | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 65 VTGN-GFEITQADGVSPTNGAPKSYPSYVDGCH-YGNCAPTTL-----PM 108
   || | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 418 PSSSPSSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSP 466
   || | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 109 RISSIGSAPS-----SVSYRYTGVYNAAYDIWL 138
   || | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 467 DSAQPDNOIKPGLQLVNTGSSVDLSTVTVRYWFTRDGSGSTLVYVNCDAAMCGGNIRAS 526
   || | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 139 DPTPTNGV-----NRTEIMWFR----- 158
   || | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 527 FGSVNPATPTADTYLQLSFTGTLAAGSGTGEIQNRVKNKSDMSNFDNDSYGTNTAF 585
   || | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 159 ---VGPVQPIGSPV-----GTAHVCGRSWEV-----WTGNSGNDVISFLAPSAI 200
   || | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 586 QDWT-KVTYVYNGRLVWG-TEPSTGTPSPSTPSPSTPSPSTPSPSTPSPSTPSPSPSPSP 643
   || | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 201 SWSFVDKDFVDQAVSHGLATPDWYLTISIQAGFEPMEGCTGLAVNSFSASVAGCGNCT 260
   || | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 644 P-TPSPSPSPSPSPSPSPSSVSSGVCGRATYVYVNSDMGSGFTATVTVNTGSRATSGMTVANSF 702
   || | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 261 PGTA-----ACQVSYSTHT-WPCGFTVDITITNTGSTPVDGWELDTL 303
   || | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 703 GGNQTVYWNALTQSGASVTATNLSYN-NVLPQOSTTTFGNGSYSGTN-TAPT----- 756
   || | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 304 PAGHTVTSVMNALISPASGAVTARSTGSGNRIANGGTQSPFGCTSSCAGTAPAGARL 363
   || | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 757 --LTCT 760
   ||
Db 364 NGTSCT 369
   ||

RESULT 11
US-09-795-583-1
; Sequence 1, Application US/09795583
; Patent No. US20020081702A1
; GENERAL INFORMATION:
; APPLICANT: Genencor International, Inc.
; TITLE OF INVENTION: No. US20020081702A1 Cellulase Producing Actinomycetes,
; FILE REFERENCE: GC540-2
; CURRENT APPLICATION NUMBER: US/09/795,583
; CURRENT FILING DATE: 2001-02-27
; PRIOR APPLICATION NUMBER: 09/321,981
; PRIOR FILING DATE: 1999-05-28
; PRIOR APPLICATION NUMBER: PCT/US99/11971
; PRIOR FILING DATE: 1999-05-28
; PRIOR APPLICATION NUMBER: 09/104,308
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 08/974,042
; PRIOR FILING DATE: 1997-11-19
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 1
; LENGTH: 371
```


TYPE: PRT
ORGANISM: Unknown
FEATURE:
OTHER INFORMATION: Nearest "neighbor" - Streptomyces
OTHER INFORMATION: thermobaceous
US-09-795-583-1

Query Match 4.8%: Score 199.5; DB 10; Length 371;
Best Local Similarity 20.8%; Pred. No. 0.00021;
Matches 101; Conservative 49; Mismatches 153; Indels 183; Gaps 22;

QY 303 FCTYHLYPNWQONADWGTOWIKDHIAAAIGKPTILEEFGWOTPDSDSVYOTWTQTVR 362
DB 39 YGTTTIQDRYVYVQNNRWGT-----SATQCIN 64
QY 363 TNGEAGNFWMLAGNV--NGQP--YPN-YDGFNVYPPSSATVLALEALAIATGTSPPPS 417
DB 65 VTGN-GFEITQADGVSPTNGAPKSPSVYDGCII-YGNCAPRTTL-----PM 108
QY 418 PSSSPSSP 466
DB 109 RISSIGSAPS-----SVSYRYTGNVYNAAYDIWL 138
QY 467 DSAPGDNQIKPGLQLVNTGSSVDLSTVTVRYWFTTRDGSSTLVYNCDAAMCGNIRAS 526
DB 139 DPTPTNGV-----NRTEIMWFNR-----158
QY 527 FCSVNPATPTADTYLQLSFTGCTLAAGSGTGEIQNRVKNKSDNSFDETND--YSYGTNTAF 585
DB 159 ---VGPVQIGSPV-----GTAHVGGRSWEV-----WTGSGNSNDV1SFLAPS 200
QY 586 QDWT-KVTYVNGRLVMG--TEPSGTSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSP 643
DB 201 SWSFVDFKDFVQAVSHGLATPDWYLTSLQAGFEPWEGGTGLAVNSFSSAVNAGGNGCT 260
QY 644 P-TPSP 702
DB 261 PCTPA-----ACQVSYSTHT-WPGGFTVDTTITNTGSTPVDGWELDFTL 303
QY 703 GGNQTVNTWNTALTQSGASVTATNLNSYN-NVLPQCSSTTFGNGSYSGTN-TAPT---- 756
DB 304 PAGHTVTSVWNLALISPASCAVTARSTGNSGRIAAAGGTQSGFQGTSSGAGTAPAGARL 363
QY 757 --LTCT 760
DB 364 NCTSCT 369

RESULT 12
US-09-739-861A-5
Sequence 5, Application US/09739861A
Patent No. US20020076792A1
GENERAL INFORMATION:
APPLICANT: Jones, Brien E.
APPLICANT: van der Kleij, Wilhelmus A. H.
APPLICANT: van Solingen, Piet
APPLICANT: Weyler, Walter
APPLICANT: Goedegebuur, Frits
TITLE OF INVENTION: No. US20020076792A1 Cellulase Producing Actinomycetes,
TITLE OF INVENTION: Cellulase Produced Therefrom and Method of Producing Same
FILE REFERENCE: GC540-3
CURRENT APPLICATION NUMBER: US/09/739,861A
CURRENT FILING DATE: 2000-12-18
PRIOR APPLICATION NUMBER: US 09/321,981
PRIOR FILING DATE: 1999-05-28
PRIOR APPLICATION NUMBER: US 09/104,308
PRIOR FILING DATE: 1998-06-24
PRIOR APPLICATION NUMBER: US 08/974,042
PRIOR FILING DATE: 1997-11-19
NUMBER OF SEQ ID NOS: 5
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 5
LENGTH: 386
TYPE: PRT
ORGANISM: Unknown

TYPE: PRT
ORGANISM: Unknown
FEATURE:
OTHER INFORMATION: Nearest "neighbor" - Streptomyces
OTHER INFORMATION: thermobaceous
US-09-739-861A-5

Query Match 4.8%: Score 198.5; DB 10; Length 386;
Best Local Similarity 20.8%; Pred. No. 0.00025;
Matches 101; Conservative 49; Mismatches 153; Indels 183; Gaps 22;

QY 303 FCTYHLYPNWQONADWGTOWIKDHIAAAIGKPTILEEFGWOTPDSDSVYOTWTQTVR 362
DB 54 YGTTTIQDRYVYVQNNRWGT-----SATQCIN 79
QY 363 TNGEAGNFWMLAGNV--NGQP--YPN-YDGFNVYPPSSATVLALEALAIATGTSPPPS 417
DB 80 VTGN-GFEITQADGVSPTNGAPKSPSVYDGCII-YGNCAPRTTL-----PM 123
QY 418 PSSSPSSP 466
DB 124 RISSIGSAPS-----SVSYRYTGNVYNAAYDIWL 153
QY 467 DSAPGDNQIKPGLQLVNTGSSVDLSTVTVRYWFTTRDGSSTLVYNCDAAMCGNIRAS 526
DB 154 DPTPTNGV-----NRTEIMWFNR-----173
QY 527 FCSVNPATPTADTYLQLSFTGCTLAAGSGTGEIQNRVKNKSDNSFDETND--YSYGTNTAF 585
DB 174 ---VGPVQIGSPV-----GTAHVGGRSWEV-----WTGSGNSNDV1SFLAPS 215
QY 586 QDWT-KVTYVNGRLVMG--TEPSGTSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSP 643
DB 216 SWSFVDFKDFVQAVSHGLATPDWYLTSLQAGFEPWEGGTGLAVNSFSSAVNAGGNGCT 275
QY 644 P-TPSP 702
DB 276 PCTPA-----ACQVSYSTHT-WPGGFTVDTTITNTGSTPVDGWELDFTL 318
QY 703 GGNQTVNTWNTALTQSGASVTATNLNSYN-NVLPQCSSTTFGNGSYSGTN-TAPT---- 756
DB 319 PAGHTVTSVWNLALISPASCAVTARSTGNSGRIAAAGGTQSGFQGTSSGAGTAPAGARL 378
QY 757 --LTCT 760
DB 379 NCTSCT 384

RESULT 13
US-09-795-583-5
Sequence 5, Application US/09795583
Patent No. US20020081702A1
GENERAL INFORMATION:
APPLICANT: Genencor International, Inc.
TITLE OF INVENTION: No. US20020081702A1 Cellulase Producing Actinomycetes,
TITLE OF INVENTION: Cellulase Produced Therefrom and Method of Producing Same
FILE REFERENCE: GC540-2
CURRENT APPLICATION NUMBER: US/09/795,583
CURRENT FILING DATE: 2001-02-27
PRIOR APPLICATION NUMBER: 09/321,981
PRIOR FILING DATE: 1999-05-28
PRIOR APPLICATION NUMBER: PCT/US99/11971
PRIOR FILING DATE: 1999-05-28
PRIOR APPLICATION NUMBER: 09/104,308
PRIOR FILING DATE: 1998-06-24
PRIOR APPLICATION NUMBER: 08/974,042
PRIOR FILING DATE: 1997-11-19
NUMBER OF SEQ ID NOS: 5
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 5
LENGTH: 386
TYPE: PRT
ORGANISM: Unknown

[illegible]

; APPLICANT: Gurney,Austin L.
; APPLICANT: Pan,James
; APPLICANT: Smith,Victoria
; APPLICANT: Matanabe,Colin K.
; APPLICANT: Wood,William I.
; APPLICANT: Zhang,Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; TITLE OF INVENTION: ACIDS ENCODING THE SAME
; FILE REFERENCE: P3430R1C1
; CURRENT APPLICATION NUMBER: US/10/052,586
; CURRENT FILING DATE: 2002-01-15
; PRIOR APPLICATION NUMBER: 60/059263
; PRIOR FILING DATE: 1997-09-18
; PRIOR APPLICATION NUMBER: 60/059266
; PRIOR FILING DATE: 1997-09-18
; PRIOR APPLICATION NUMBER: 60/062250
; PRIOR FILING DATE: 1997-10-17
; PRIOR APPLICATION NUMBER: 60/063120
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: 60/063121
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: 60/063486
; PRIOR FILING DATE: 1997-10-21
; PRIOR APPLICATION NUMBER: 60/063540
; PRIOR FILING DATE: 1997-10-28
; PRIOR APPLICATION NUMBER: 60/063541
; PRIOR FILING DATE: 1997-10-28
; PRIOR APPLICATION NUMBER: 60/063544
; PRIOR FILING DATE: 1997-10-28
; PRIOR APPLICATION NUMBER: 60/063564
; PRIOR FILING DATE: 1997-10-28
; PRIOR APPLICATION NUMBER: 60/063734
; PRIOR FILING DATE: 1997-10-29
; PRIOR APPLICATION NUMBER: 60/063870
; PRIOR FILING DATE: 1997-10-31
; PRIOR APPLICATION NUMBER: 60/064103
; PRIOR FILING DATE: 1997-10-31
; PRIOR APPLICATION NUMBER: 60/065311
; PRIOR FILING DATE: 1997-11-13
; PRIOR APPLICATION NUMBER: 60/066120
; PRIOR FILING DATE: 1997-11-21
; PRIOR APPLICATION NUMBER: 60/066466
; PRIOR FILING DATE: 1997-11-24
; PRIOR APPLICATION NUMBER: 60/066772
; PRIOR FILING DATE: 1997-11-24
; PRIOR APPLICATION NUMBER: 60/069335
; PRIOR FILING DATE: 1997-12-11
; PRIOR APPLICATION NUMBER: 60/069425
; PRIOR FILING DATE: 1997-12-12
; PRIOR APPLICATION NUMBER: 60/069870
; PRIOR FILING DATE: 1997-12-17
; PRIOR APPLICATION NUMBER: 60/068017
; PRIOR FILING DATE: 1997-12-18
; PRIOR APPLICATION NUMBER: 60/077450
; PRIOR FILING DATE: 1998-03-10
; PRIOR APPLICATION NUMBER: 60/077632
; PRIOR FILING DATE: 1998-03-11
; PRIOR APPLICATION NUMBER: 60/077649
; PRIOR FILING DATE: 1998-03-11
; PRIOR APPLICATION NUMBER: 60/078886
; PRIOR FILING DATE: 1998-03-20
; PRIOR APPLICATION NUMBER: 60/078939
; PRIOR FILING DATE: 1998-03-20
; PRIOR APPLICATION NUMBER: 60/079664
; PRIOR FILING DATE: 1998-03-27
; PRIOR APPLICATION NUMBER: 60/079786
; PRIOR FILING DATE: 1998-03-27
; PRIOR APPLICATION NUMBER: 60/080107
; PRIOR FILING DATE: 1998-03-31
; PRIOR APPLICATION NUMBER: 60/080194
; PRIOR FILING DATE: 1998-03-31
; PRIOR APPLICATION NUMBER: 60/080327
; PRIOR FILING DATE: 1998-04-01
; PRIOR APPLICATION NUMBER: 60/080333
; PRIOR FILING DATE: 1998-04-01
; PRIOR APPLICATION NUMBER: 60/081049
; PRIOR FILING DATE: 1998-04-08
; PRIOR APPLICATION NUMBER: 60/081070
; PRIOR FILING DATE: 1998-04-08
; PRIOR APPLICATION NUMBER: 60/081195
; PRIOR FILING DATE: 1998-04-09
; PRIOR APPLICATION NUMBER: 60/081838
; PRIOR FILING DATE: 1998-04-15
; PRIOR APPLICATION NUMBER: 60/082568
; PRIOR FILING DATE: 1998-04-21
; PRIOR APPLICATION NUMBER: 60/082569
; PRIOR FILING DATE: 1998-04-21
; PRIOR APPLICATION NUMBER: 60/082704
; PRIOR FILING DATE: 1998-04-22
; PRIOR APPLICATION NUMBER: 60/082797
; PRIOR FILING DATE: 1998-04-22
; PRIOR APPLICATION NUMBER: 60/083322
; PRIOR FILING DATE: 1998-04-28
; PRIOR APPLICATION NUMBER: 60/083495
; PRIOR FILING DATE: 1998-04-29
; PRIOR APPLICATION NUMBER: 60/083496
; PRIOR FILING DATE: 1998-04-29
; PRIOR APPLICATION NUMBER: 60/083499
; PRIOR FILING DATE: 1998-04-29
; PRIOR APPLICATION NUMBER: 60/083559
; PRIOR FILING DATE: 1998-04-29
; PRIOR APPLICATION NUMBER: 60/084366
; PRIOR FILING DATE: 1998-05-05
; PRIOR APPLICATION NUMBER: 60/084414
; PRIOR FILING DATE: 1998-05-06
; PRIOR APPLICATION NUMBER: 60/084639
; PRIOR FILING DATE: 1998-05-07
; PRIOR APPLICATION NUMBER: 60/084640
; PRIOR FILING DATE: 1998-05-07
; PRIOR APPLICATION NUMBER: 60/084643
; PRIOR FILING DATE: 1998-05-07
; PRIOR APPLICATION NUMBER: 60/085573
; PRIOR FILING DATE: 1998-05-15
; PRIOR APPLICATION NUMBER: 60/085579
; PRIOR FILING DATE: 1998-05-15
; PRIOR APPLICATION NUMBER: 60/085580
; PRIOR FILING DATE: 1998-05-15
; PRIOR APPLICATION NUMBER: 60/085582
; PRIOR FILING DATE: 1998-05-15
; PRIOR APPLICATION NUMBER: 60/085700
; PRIOR FILING DATE: 1998-05-15
; PRIOR APPLICATION NUMBER: 60/086023
; PRIOR FILING DATE: 1998-05-18
; PRIOR APPLICATION NUMBER: 60/086392
; PRIOR FILING DATE: 1998-05-22
; PRIOR APPLICATION NUMBER: 60/086486
; PRIOR FILING DATE: 1998-05-22
; PRIOR APPLICATION NUMBER: 60/087098
; PRIOR FILING DATE: 1998-05-28
; PRIOR APPLICATION NUMBER: 60/087208
; PRIOR FILING DATE: 1998-05-28
; PRIOR APPLICATION NUMBER: 60/087609
; PRIOR FILING DATE: 1998-06-02
; PRIOR APPLICATION NUMBER: 60/087759
; PRIOR FILING DATE: 1998-06-02
; PRIOR APPLICATION NUMBER: 60/087827
; PRIOR FILING DATE: 1998-06-03
; PRIOR APPLICATION NUMBER: 60/088025
; PRIOR FILING DATE: 1998-06-04
; PRIOR APPLICATION NUMBER: 60/088028
; PRIOR FILING DATE: 1998-06-04
; PRIOR APPLICATION NUMBER: 60/088029
; PRIOR FILING DATE: 1998-06-04
; PRIOR APPLICATION NUMBER: 60/088033
; PRIOR FILING DATE: 1998-06-04
; PRIOR APPLICATION NUMBER: 60/088167

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM nucleic - nucleic search, using sw model
Run on: November 13, 2002, 08:50:51 ; Search time 3966 Seconds
(without alignments)
16796.854 Million cell updates/sec
Title: US-09-917-378-2
Perfect score: 2289
Sequence: 1 atgggtcctagtcgctgcgc.....tcacctgcacggctagtlga 2289
Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0
Searched: 2054640 seqs, 14551402878 residues
Total number of hits satisfying chosen parameters: 4109280
Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Result No.	Score	Query Match %	Length	DB	ID	Description
1	205	9.0	3237	1	AF200304	AF200304 Caldicell
2	180.4	7.9	3004	1	ACU33212	U33212 Acidithiobac
3	180.4	7.9	3004	6	AX467594	AX467594 Sequence
4	180.4	7.9	3004	6	123701	123701 Sequence 6
5	172.8	7.5	4567	1	AF163837	AF163837 Caldicell
6	133.6	5.8	11707	1	AF078737	AF078737 Caldicell
7	133.6	5.8	11707	6	E35099	E35099 Truncated c
8	128.4	5.6	6005	1	AF078042	AF078042 Caldicell
9	128.4	5.6	6416	6	E35100	E35100 Truncated c
10	126.4	5.5	1992	6	AR205120	AR205120 Sequence
11	126.4	5.5	2043	1	TMA17980	Y17980 Thermotoga
12	126.4	5.5	2043	6	AR205088	AR205088 Sequence
13	124.8	5.5	20990	1	AE001779	AE001779 Thermotog
14	124.4	5.4	2001	1	AY033477	AY033477 Thermotog
15	124.4	5.4	2001	1	TNE17981	Y17981 Thermotoga
16	122	5.3	5437	1	CASR69XYN2	AF036924 Caldicell
17	118.6	5.2	1265	8	ASNMANNA	L35487 Aspergillus
18	118.6	5.2	1302	6	AR023984	AR023984 Sequence
19	116.4	5.1	4977	1	CDCMANA	L01257 Caldocellum
20	114.8	5.0	2977	6	A28170	A28170 B. laetus st
21	113.2	4.9	2831	1	BACCELA	M76588 Bacillus la
22	110.6	4.8	7686	1	AF038547	AF038547 Bacillus
23	106.2	4.6	5439	1	CDCCELA	L32742 Caldocellum
24	102.8	4.5	4743	1	AB008029	AB008029 Thermophi
25	100.8	4.4	5513	1	AT286105	286105 A. thermophi
26	98.4	4.3	3262	1	AF078038S1	AF078038 Caldicell
27	98.2	4.3	5284	1	CSU16708	U16708 Caldocellum
28	97.8	4.3	1621	1	THFE2AA	M73321 Thermomonas
29	97.8	4.3	3509	1	PAE488923	AJ488933 Paenibac
30	92.2	4.0	1434	8	ABCELA4MR	250095 Aquarius bi
31	90.2	3.9	2029	6	E35142	E35142 Truncated c
32	90.2	3.9	2029	6	E35143	E35143 Truncated c
33	88.4	3.9	210319	2	AL607086	AL607086 Mus muscu
34	87.8	3.8	40104	1	SCF11	AL132662 Streptomy
35	87	3.8	165165	2	AC096270	AC096270 Rattus no
36	86.4	3.8	37941	1	AOPKZA361	AJ223998 Amycolato
37	86.4	3.8	41906	1	SC5C7	AL031515 Streptomy
38	85.8	3.7	4161	1	BSP133614	AJ133614 Bacillus
39	85.4	3.7	3480	1	CFICEND	L02544 Cellulomona
40	85.2	3.7	2801	5	AF218784	AF218784 Gallus ga
41	85.2	3.7	4241	1	CSCELB	X13602 Caldocellum
42	84.6	3.7	163636	10	AL669937	AL669937 Mouse DNA
43	84	3.7	207070	10	AC122843	AC122843 Mus muscu
44	83.8	3.7	185241	2	AC115962	AC115962 Mus muscu
45	83.6	3.7	108360	14	HSTULR	D10875 Herpes simp

ALIGNMENTS

RESULT 1
AF200304
LOCUS
DEFINITION
AF200304
3237 bp DNA linear BCT 01-DEC-2000
Caldibacillus cellulovorans putative type IIb cellulose-binding
domain gene, partial cds; and beta-1,4-xylanase xynA precursor
(xynA) gene, complete cds.
ACCESSION
AF200304
VERSION
AF200304.1 GI:7385018
KEYWORDS
Caldibacillus cellulovorans.
SOURCE
Caldibacillus cellulovorans
ORGANISM
Bacteria; Firmicutes; Bacillales; Alicyclobacillaceae;
Caldibacillus.
REFERENCE
1 (bases 1 to 3237)
AUTHORS
Sunna,A., Gibbs,M.D. and Bergquist,P.L.

Pred. No. is the number of results predicted by chance to have a


```

/misc_feature
/ gene="manA"
/product="multidomain beta-1,4-mannanase"
1349..11480
/ gene="manA"
/ note="Region: proline-threonine linker"
/ evidence=not_experimental
1481..11927
/misc_feature
/ gene="manA"
/ note="Region: cellulose-binding domain type 111b"
/ evidence=not_experimental
1928..2032
/misc_feature
/ gene="manA"
/ note="Region: proline-threonine linker"
/ evidence=not_experimental
2033..2884
/misc_feature
/ gene="manA"
/ note="Region: beta-1,4-mannanase catalytic domain"
/ evidence=experimental
2885..2959
/misc_feature
/ gene="manA"
/ note="Region: proline-threonine linker"
/ evidence=not_experimental
2960..3409
/misc_feature
/ gene="manA"
/ note="Region: cellulose-binding domain type 111b"
/ evidence=not_experimental
3894..4400
/ note="orf3"
/ codon_start=1
/ transl_table=11
/ product="unknown"
/ protein_id="AAF22275.1"
/ db_xref="GI:6651328"
/ translation="MGASVFSRTAONGYGLKARQKRCVGRNLLGHRI.FVVDDEF
RFFACARLRFRSFRONRLCAQNFESTVTFVQDDKRRIRRENICQLFVRFQASI
GPKRLIFIAEQWQARKQASIQPOLKNAVLIFIDQHMPYGVEMDPFHKEFKQS
RSILRTY"
HASE COUNT 1040 a 1276 c 1439 g 812 t
ORIGIN

Query Match 7.5%; Score 172.8; DB 1; Length 4567;
Rest Local Similarity 55.8%; Pred. No. 3.4e-14;
Matches 393; Conservative 0; Mismatches 302; Indels 9; Gaps 3:

Qy 1241 CTCCGCGCTCGCGAGCTCGAGTCCATCTCGTCGCGCTCTCGTCGCGCTCTCCGTCGG 1300
Db 1350 CCGCGCGCGCTCGCGACGACCGCGACACCGGACCGGACACCGTCCGACGC 1409
Qy 1301 CGTCTCGTCCGCGCTCTCGTCGCGCTCTCTGTCGCGGAGCGCGTCTCGTCTCGCG 1360
Db 1410 CGACACCGGCGCCACCGAGTGTGACACCAACCGTGCACCGGCTCGACCGCGACGC 1469
Qy 1361 CGCTGTCTCGGCTGGG---TGAAGGTGCAGTACAAGAACAAATATTCCGCGCGGTGATA 1417
: 1373 TAGCGGAAGCGCACCTCTCGCGTGCAGTATCGCTGGCGGACACAGCGGCGACCGACA 1529
1418 ACCAGATCAACCGGCTCTCCAGTTGGTGAATACGGGGTCTGTCGTCGGTGGATTTGTCGA 1477
Db 1530 ACCAGATCAACCGGAGCTCGCATCTGTCACACCGCGCTCGAAGCGCTGCCGCTCAGCC 1589
Qy 1478 CGGTACCGTGGCGTACTGTTTACCCCGGATGGTGGTCTGTCGACACTGGTGTACAAC 1537
Db 1590 AGCTGAAGTGGCGTACTCGTACAC---GAAGAACTCGACGCGGCGGCAACAGTACTTCT 1646
Qy 1538 GTGACTGGCGCGGCTGGGTGTGGGAATATCCGGCGCTCTGTTGGGCTCGGTGAACCGG 1597
Db 1647 GCGACTGGCGCGAGATCGGCTCTCGAACATCCGGCGGCGAGTTCGTCTCGTGTCCGACC 1706
Qy 1598 CGACCGCGGCGGACACCTACCTCGAGTTGCTGTTCACTGGTGAACGTTGGCGGCTG 1657
Db 1707 CGGTACGCGGCGGCGGACGATCATCTGAGTCTACGAGTTCACGGCGGCAAGCAITTCGCGCG 1766
Qy 1658 GTGGGTGCGAGGTCAGATTCAAAACCGCGGTCAATAAGAGCTACTGGTCCAACTTTTGATG 1717

```

```

Db 1767 GAGGCAACCGCGGAGATTACAGACCGGATTCACTTTCAGCAACTGCATCAACTACCAACG 1826
Qy 1718 AGACCAATGACTACTCGT---ATCGGACGGAACACCGCTTCCAGGATTGGACGAAAGCTGA 1774
Db 1827 AACGGACGACTCGTGTACACGGGCGGACAGCGTGGGGCGGCTCGACCGCGGATTA 1886
Qy 1775 CGGTGTATGTCATGCGCGCTGCTGGGGGACTGAACCGTCCGCGCACCAACCGCCACG 1814
Db 1887 CGCTTTATCCACACGGGCTGCTGTATGGGGGAGGAGCGCGGCGGCGGATCTGTCGCGG 1946
Qy 1835 CCACACCGACCGCCAGCCCAACCGCTGCCCGGACCGCGGACCGCGGACCGCCAGCT 1894
Db 1947 CGACCGCGAGGTGACACCGACACCGGACCGCGGACGTCAACCGCGGACACCTACG 2006
Qy 1895 CCTCCCATCCCGTCCCGGAGCGCCGACGCGCGGACCGCTAGCGCG 1938
Db 2007 CTACGCCAATCAACGCCGACACCGGAGCGGTGACCGGAACCTG 2050

RESULT 6
AF078737
LOCUS
DEFINITION
AF078737 1107 bp DNA linear BCT 31-JAN-2001
complete cds; and unknown genes.
ACCESSION
AF078737
VERSION
AF078737.1 GI:4836163
KEYWORDS
SOURCE
ORGANISM
Caldicellulosiruptor sp. Tok7B.1.
Caldicellulosiruptor sp. Tok7B.1
Bacteria; Firmicutes; Clostridia; Clostridiales;
Syntrophomonadaceae; Caldicellulosiruptor.
REFERENCE
1 (bases 1 to 1107)
Gibbs,M.D., Reeves,R.A., Farrington,G.K., Anderson,P.,
Williams,D.P. and Berquist,P.L.
Multidomain and multifunctional glycosyl hydrolases from the
extreme thermophile Caldicellulosiruptor isolate Tok7B.1
Curr. Microbiol. 40 (5), 333-340 (2000)
20171169
10706665
PUBMED
2 (bases 1 to 1107)
Gibbs,M.D., Reeves,R.A., Farrington,K.G., Anderson,P.,
Williams,D.P. and Berquist,P.L.
Direct Submission
Submitted (17-JUL-1998) School of Biological Sciences, Macquarie
University, Sydney, New South Wales 2109, Australia
FEATURES
Location/Qualifiers
1..1107
/organism="Caldicellulosiruptor sp. Tok7B.1"
/strain="Tok7B.1"
/db_xref="taxon:80339"
<1..683
/ note="orf1"
/ codon_start=3
/ transl_table=11
/ product="unknown"
/ protein_id="AAD30365.1"
/ db_xref="GI:4836169"
/ translation="NLCHVLKORFAVHTRCAHKKIKYRNI SPALSSPPSSISKPFQ
DSGLSYASLASAPAI VDDVRSILTFWQDSILSKKDEIRNMYGDMKKPPAPQVAVG
PAEYKWTATQINDSDVNSNLIPPLQSGDSLVLMTQTDMSPSGNVIRNGVFI SLA
EYTFGNVNSGDLKIWDPRSOOTINEITNDLNLPIVPTPTVPPTANVTGTTNNFQI
INRRNSIE"
729..6041
/ gene="xynA"
729..6041
/ gene="xynA"
/ note="xylanase/arabinosidase: multidomain enzyme with an
amino-terminal glycosyl hydrolase family 10 domain, and a
carboxy-terminal glycosyl hydrolase family 43 domain"
/ codon_start=1
/ transl_table=11
/ product="xynA"

```


VERSION E35099.1 GI:13018924
KEYWORDS JP 1999221086-A/1.
SOURCE unidentified.
ORGANISM unclassified.
REFERENCE 1 (bases 1 to 11707)
AUTHORS Palji.A., Petaer,L.B., Koy,M.D., Gurahamu,K.F., Moreland,D.G.,
Hyu,M. and Baian,P.W.
TITLE Truncated cellulase composition
JOURNAL Patent: JP 199221086-A 1 17-AUG-1999;
CLARIANT INTERNATIONAL LTD
COMMENT OS Unknown
PN JP 1999221086-A/1
PD 17-AUG-1999
PF 21-SEP-1998 JP 1998283606
PR 19-SEP-1997 US 08/932571
PI PALJI ANDERSON,PETAER L BAGUKUISUTO,KOY M DANIEL, PI
GURAHAMU K PARINTON,
PI MORELAND DAVID GIBUSU,HYU MORGAN,DAIAN PURATONIORISU WILLIAM
PC C12N15/09,C11D3/386,C12N1/21,C12N9/42//((C12N1/21,C12R1:19), PC
(C12N9/42,C12R1:19),C12N15/00
CC Key Location/Qualifiers
FH source 1..11707 /organism='Unknown'.
FT
FT
FEATURES
source
1..11707
/organism='unidentified'
/db_xref='taxon:32644'
BASE COUNT 3828 a 1994 c 2994 g 2891 t
ORIGIN
Query Match 5.8%; Score 133.6; DB 6; Length 11707;
Best Local Similarity 52.4%; Pred. No. 5.3e-09;
Matches 395; Conservative 0; Mismatches 344; Indels 15; Gaps 4;
QY 1220 CAATCAGTACCGGCACATCGCCTCCCGCTCCCGAGCTCGAGTCATCCTCGTCCGCGT 1279
DB 3883 CAACCAACACCTCGACCGACATCAGCCCCGACACAGCCAGCCACAGTACAGCAAGCCGCA 3942
QY 1280 CTCCTCGCGCTCTCGCTCGCGGCTCTCCGCTCGCGCTCTCCGCTCGCGCTCTCGTCCGCCA 1339
DB 3943 CTCCAACCGCGACCGACAGTACGCTTCTGCTACTCCGACCGACACCAACACCCCA 4002
QY 1340 GCGCTCTCCGCTCGCTCGCGCGGCTGTGCGG---GTGGGCTGAAGTGCAGTACAAGACA 1396
DB 4003 CGCCGACACCGACAGGACACCTGGCAGCGGAAGTGGTTGAAGGTACTATACAAGACA 4062
QY 1397 ATGATTCCGCGCGCGGTGATACACAGATCAACCGCGCTCCAGTTGGTGAATACGGGT 1456
DB 4063 ATGACAGAAGTCCGAGACACAGTCTTATAAGCCCGTGGTTTAAGATAGTGAATGGAGCA 4122
QY 1457 CUTCCTCGCTGGATTGTTCGACGGTGCAGCTGCGGTGCTGCTTACCCCGGATGGTGGT 1516
DB 4123 CGACGAGTGTGATCTTACGAGGTTAAGATAAGATACTGTTACAGAGTGGATGGTG--- 4179
QY 1517 GCTCAGACTGGTGTACAACTGTACTCGCGCGCGATGGGTGTGGGAATATCCCGCGCT 1576
DB 4180 ACAAGCCACAGAGTGGCTGTGTACTGGGACAGATAGGGGCAAGCAATGTGACATTCA 4239
QY 1577 CTTCCGCTCGGTGAACCGCGCGACCGCGGACACCTTACCTGCAGTGTCTGTCA 1636
DB 4240 ATTTTGTGAAGCTGACGAGCGAGTGAAGTGGAGCGGATTTACTTGGAGGTAGGATTTA 4299
QY 1637 CTGTTGGAAAC-----GTTGGCGCTGGTGGGTGCGAGCGGTGAGATTCAAAACCGGTGA 1690
DB 4300 CGAGTGGAGCTGGGCAGTTTCAGGCGCTGTGAAGACACAGGGATATACAGGTAAGGTTTA 4359
QY 1691 ATAAGAGTACTGGTGAACCTTTGATGAGACCAATGACTACTCTGTATGGGAGCAACCG 1750
DB 4360 ACAAGAACTGACTGGGAACTTACAATCAGGACGACCTGCTGTTGACAGACATGA 4419
QY 1751 CTTCCAGGATTGGA---CGAAGGTGACGGTCTATGTCAATCGCCGCGCTGGTCTCGGGCA 1807

DB 4420 CGAATTTATCGAGAGAAATCGAAGTTAAGCTGTATGTATAGATGCTGTCTGCTATGGGCG 4479
QY 1808 CTGAACCGTCCGGCAGCAGCCCGCCAGCCACACCCAGCCCGCCAGCCCGCTCCCGCA 1867
DB 4480 AGGAGCCGGCGGAGCGACACCTTGCACGACAGCAACCAACCCCAACTCCGA 4539
QY 1868 GCCCGAGCCGACCCCAAGCCAGCTCTTCCCATCTCCCGTCCCGGAGCCCGAGCCCA 1927
DB 4540 CAGCAACACAGACAGCCGCGACAGCAAGCGCAAGCTCTACACCGACACCGACACA 4599
QY 1928 GCCCTACGCGTCCCGTCCCGAGCCGTCGCGC 1961
DB 4600 CACCAACCCCAATACCAACACCCACACGCGCTCC 4633
RESULT 8
AF078038S5 6005 bp DNA linear HCT 11-FEB-2001
LOCUS Caldicellulosiruptor sp. Tok7B.1 glycosyl hydrolase 6 genes, partial
DEFINITION cds; and Cele gene, complete cds.
ACCESSION AF078042
VERSION AF078042.1 GI:12743878
KEYWORDS
SEGMENT
SOURCE
ORGANISM
5 of 5
Caldicellulosiruptor sp. Tok7B.1.
Caldicellulosiruptor sp. Tok7B.1
Bacteria; Firmicutes; Clostridia; Clostridiales;
Syntrophomonadaceae; Caldicellulosiruptor.
REFERENCE 1 (bases 1 to 6005)
AUTHORS Gibbs,M.D., Reeves,R.A., Farrington,G.K., Anderson,P.,
Williams,D.P. and Bergquist,P.L.
TITLE Multidomain and multifunctional glycosyl hydrolases from the
extreme thermophile Caldicellulosiruptor isolate Tok7B.1
JOURNAL Curr. Microbiol. 40 (5), 333-340 (2000)
MEDLINE 20171169
PUBMED 10706665
REFERENCE 2 (bases 1 to 6005)
AUTHORS Gibbs,M.D., Reeves,R.A., Farrington,G.K., Anderson,P.,
Williams,D.P. and Bergquist,P.L.
TITLE Direct Submission
JOURNAL Submitted (27-OCT-1999) Biological Sciences, Macquarie University,
Sydney, NSW 2109, Australia
FEATURES
Location/Qualifiers
source
1..6005
/organism='Caldicellulosiruptor sp. Tok7B.1'
/isolate='Tok7B.1'
/db_xref='taxon:80339'
CDS
1..506
/note='ORF6; multidomain'
/codon_start=3
/transl_table=11
/product='glycosyl hydrolase 6'
/protein_id='AAK06393.1'
/db_xref='GI:12743884'
/translation='GQPDITWNTYTGPNLHKV1VDYGTDLGTLASLANALLYSAAT
KEYGVSDEAAKNLAKELDRMNLVGRDGLSLAPEKRGDYKRFFQEVYIPAGWTGKM
PNGDVIKSGVKFIDIRSKYKODPDQKQLVSAYNAGEAEFRHFRWAQCDIAIANTY
EILFQK'
misc_feature
1..503
/note='Region: family 48 glycosyl hydrolase domain'
CDS
617..5872
/note='multidomain glycosyl hydrolase'
/codon_start=1
/evidence=experimental
/transl_table=11
/product='Cele'
/protein_id='AAK06394.1'
/db_xref='GI:12743885'
/translation='MOENKAIKRYYVITALLVTLISLCPFGIMPVKAYAGTYNTYGEA
LQRTIMFYFQNSKLPWSVRNMRGSDGLDDGLDLTGGWHADGDKVAFNLPMS
YASMLGWAVYKDAFYKSKOLEHILNQIEWANDYFYKCHPSKYVYVYQVGDPTVDH
NFWGPAEYVHMKRPAYKCDLSNPASSVVAETAASLAVASVVIKERNOKAASYLQHAK

Db	2593	ACACCAACAGTGCACAGCAACCCGACGCGCGACCTCCCTACACCGACACCGACCGGGTACCT	2652
Qy	1363	GTGTGCG--GTGGGGTCAAGGTGCAGTACAGAACAATGATTTCGGCGCGGTGNTAAC	1419
Db	2653	GCTACGGGAAGTGGTGCAGGTACTGTACAGAACAATGACACAAGTCCGAGCAGACAGT	2712
Qy	1420	CAGATCAACACGGGTCTCCAGTTGGTGAATACGGGGTCTCTCGTCGGTGCATTTGTTCGACG	1479
Db	2713	TCATATAAGCCCGTGGTTTAAAGATAGTGAATGGAGCAGCAGCAGTGTGTATCTTAGCAGG	2772
Qy	1480	GTGACGGTCCGTACTGCTTACCACGGGATGGTGGTTCGTCCGACACTGGTGTACAACCTGT	1539
Db	2773	GTTAAGATAAGATAGTGTACACAGTGGATGGTG---ACAAGCCACAGAGTCCGCTATGT	2829
Qy	1540	GACTGGGCGGGATGGGTGTGGGAATATCGCGCGCTCTGTCGGCTGAACCCGCGC	1599
Db	2830	GACTGGGCACAGATAGGGGCAAGCAATGTGACATTTCAATTTTGTCAAGCTTAGCAGCGGA	2889
Qy	1600	ACGGCGACGGGGACACCTACTCTGAGTTGTGTTCACTGTGTGAAC-----GTTGCCC	1653
Db	2890	GTGAGTGGAGCGGATTTATTAAGTCTGAGGTAGGATTTAGCAGTGGAGCTGGCGAGTTCCAG	2949
Qy	1654	CTGTGTGGTTCGACGGGTGAGATTCAAAACCGGTGAATAAGACTGACTGGTCGAACCTTT	1713
Db	2950	CGTGTAGGACACAGGGGATATACAGTAAAGTTTAAACAAGATGACTGGAGCAATTAC	3009
Qy	1714	GATGAGACCAATGACTACTCTGTATGGGACGAACACCGCTTCCAGGATTGGA---CGAAG	1770
Db	3010	AATCAGCAGACGACTGCTCATGTTGTCAGAGCATGACGAATATATGGAGAGAAATCCGAAG	3069
Qy	1771	GTGACGGTCTATGTCAATTCGGCGGTGGTGTGGGGGACTGAACCGTCCGGCACCAGCCCC	1830
Db	3070	CTGACGCTCTATGTAGATGGTGTCTGCTGTGGGGCAGGAGCCGGGAGGAGCGACCT	3129
Qy	1831	AGCCCAACACCCAGCCCGACACCCGCTCCCGAGCCGAGCCGACCCCAAGACCCC	1890
Db	3130	GCACCGACAAGCAGACGAACCAACAGCCCACTCCGACGAGAACCCCAACACTACACT	3189
Qy	1891	AGCTCTCCCATCCCGTCCCGAGCCCCAGCCCCAGCCCTACGCCGTCCCGCTGGCGC	1950
Db	3190	ACACGACCCCGACCCACACAGTGAATGTCAGCGCAACCCGACCCGACCGCATCACG	3249
RESULT 10			
AR205120		1992 bp	DNA
LOCUS	Sequence 57 from patent US 6368844.	linear	PAT 20-JUN-2002
DEFINITION			
ACCESSION	AR205120		
VERSION	AR205120.1	GI:21502622	
KEYWORDS			
SOURCE	Unknown.		
ORGANISM	Unknown.		
REFERENCE			
AUTHORS	1 (bases 1 to 1992)		
TITLE	Bylina,E.J.		
JOURNAL	Glycosidase enzymes		
FEATURES	Patent: US 6368844-A 57 09-APR-2002;		
source	Location/Qualifiers		
	1..1992		
	/organism="unknown"		
BASE COUNT	600 a	405 c	573 g
ORIGIN	414 t		
Query Match	5.5%;	Score 126.4;	DB 6;
Best Local Similarity	49.3%;	Pred. No. 6.5e-08;	
Matches 513;	Conservative	0;	Mismatches 476;
		Indels 51;	Gaps 5;
Qy	121	TTCTGTACCGCATCCGGGGTTCAGTTCTTCTGAACGGCTTCCTATCGTTACGGGGA	180
Db	49	TTCTGAAAGTGGAAAACGGAAAAATTCCTCTGAACGGAAGAAATTCAGATTTCATGGA	108
Qy	181	ACGAACAACATATTACCTCAGCTATCAGTCCGACCCGACGTCGATGAGCTTTCGGCAAG	240

[illegible]

RESULT 11	DEFINITION
TMA17980	ACCESSION
LOCUS	VERSION

Accession	Size	Library	Source
TM17980	2043 bp	DNA	BCT 30-SEP-1999
Thermotoga maritima manB gene.			
Y17980			
Y17980.1	GI:6006594		


```

KEYWORDS      beta-mannosidase; manB gene.
SOURCE        Thermotoga maritima.
ORGANISM      Bacteria; Thermotogae; Thermologae; Thermologales; Thermotogaceae;
               Thermotoga.
REFERENCE     1 (bases 1 to 2043)
AUTHORS      Parker,K.N., Lam,D., Duffaud,G., Sincad,M.A., Mathur,E.J., and
               Kelly,R.M.
TITLE        Amino Acid Sequence of beta-mannosidase genes from the
               hyperthermophilic bacteria Thermotoga maritima and Thermotoga
               neapolitana
JOURNAL      Unpublished
REFERENCE     2 (bases 1 to 2043)
AUTHORS      Parker,K.N.
TITLE        Direct Submission
JOURNAL      Submitted (21-AUG-1998) K.N. Parker, North Carolina State
               university, Department of Chemical Eng. Box 7905 NCSU, Raleigh, NC
               27695-7905, USA
FEATURES     Location/Qualifiers
               1..2043
               /organism="Thermotoga maritima"
               /strain="MSB8"
               /db_xref="taxon:2336"
               1..2043
               /gene="manB"
               1..2043
               /db_xref="GI:6006595"
               /translation="MGIGDDSDSPSVSAEFLILLIVELSPVLFASDEFVKVNGKFFAL
               NCKEFRFGISNNYTHMYKNSGMDISVLESARDMGIKVLRIWGLDGSYCYDKNNTYMH
               PEPGVFGPEGISNAQSGFEKLDITVAKAKELGKLVILVNNDDFGGMQYVRWFG
               GTHHDDVYRDEK IKEEVKYVSVLVNHTVTGVPYREEPTIMAWELANEPRCETDKS
               GNTLVENWKWASSYIKSLDPNHLVAVGDEGFSSNYEGKFPYGGGAENWYNGSGVDWK
               KLLSIEVDFGTFHLYPHSHWCYSPENYAQWCAKWIEDIHRIKAEIGRPVLUVEYGIPIK
               SAPVNRATYRLMNDLVLDLGGDGMFWMLAGI GEGSDRDERGYYPDYDGRFTVNDSDS
               PEALIRYAKLFTGEDIREDTCSFLLPKDMEIKKTVREVRAGVFYDYNFTKLSVK
               VEDLVFENETBILGAYGIYGLDITTRIPDGHEMFLEGHFOGKTVKDSIKAKVYNEAR
               VYLAEDVDFSPERKNNNSQWQALFQSPDIWNGEVGNCAQLNKLKPSDWEE
               VRVAKFERLSECEILKVDIYIPNVEGLKRLRPYAVLNPQWVKIGLDMNNANVESAE
               IITPCGKEYRHFHWRJIEFDRTAGYKEHLIGVVGDLHLYDGFIFIDNVRLYRTGGM"
BASE COUNT   609 a 417 c 592 g 425 t
ORIGIN
Query Match      5.5%; Score 126.4; DB 1; Length 2043;
Best Local Similarity 49.3%; Prod. No. 6.4e-08;
Matches 513; Conservative 0; Mismatches 476; Indels 51; Gaps 5;

QY 121 TTCTGTCACCGCATCGCGGGTCAGTTCGTTTGAACGGCCCTTCCCTATCGTTACGGGGGA 180
Db 100 TTCTGTGAAGTGGAAACGGAAATTCGCTCTGAACCGGAAAGAAATTCAGATTTCATTGGA 159
QY 181 AGGAACAACATATTACCTCAGCTATCAGTTCGACACGGCGGACGTCGATTCAGCTGTTGCCAAG 240
Db 160 AGCAACAACATCTACATGCATCAGGAGCAACGGAATGATAGACAGTCTTCTCGGAGAGT 219
QY 241 GCTCAAGCGCATCAATCTTTCTCTCATCGGACCTCGGGTTCATCCACATCGGCTCTCTT 300
Db 220 CCCAGAGACATGGGTATAAAGGTCCTCAGATCTCGGGTTCCTTCGACGGGGAGACTTAC 279
QY 301 CAGCGCTCCGTCGCCACANTCGATGGCAACGAACAGCGCTTCTACTTTCAGTACTGGGAC 360
Db 280 TCCAGAGACAAGAACACCTACATGATCTCTGAGCCCGGTGTTTTCGGGTGTCAGAAGGA 339
QY 361 CCG..CGACGGCGCTTCGGCGGTACAAACACGGCGGCGGCTTCCAGGCGCTTGCAGCTAC 420
Db 340 ATATCGAAGGCC-----CAGACGGGTTCGAAAGACTTCGACTAC 378
QY 480 GCGATCGCGAGCGGGCGCGCGCGCTTCGGGTGATTTGTCGCTCCTCACCAACGACTCG 480

```

```

Db 379 ACAGTTTCCAAAGCGAAAGAACTCGGTATAAAACTTGTCTATTGTTCTTGTGAACAACG 438
QY 481 AAAGAATTTGGGGGAATGATCAATAACGAACTGCTGATCGGCCTTCTTACCACGACAAAC 540
Db 439 GACGATTCGGTGGAAATGAACCACTACGTCAGGTGGTTTGGAGGAACCCATCACGACGAT 498
QY 541 TTCTACACGACGCCCCCGGACCGAGCGGTATCAAGAAATTTGGGTCAATCATCTACTCAAC 600
Db 499 TTCTACAGAGATGAGAAGATCAAGAAGAGTACAAAAGTACGTCCTTCTTCGTGTAAC 558
QY 601 CGGTCACACGATTTACCGCGGTGACGTACAGAAGAGTCCAAACGATCTTTGTTGGGGA 660
Db 559 CATGTCAATACCTACACGGGAGTTCTTACAGGGAAGAGCCCACTCATGGCTCGGAG 618
QY 661 CTTTGCAATGAGCGCGCTCGGTAGGACGCGCACATTACCAACCTTCGGGCGACGTCGACT 720
Db 619 CTTGCAACCAACCGCGCTGTGAGCGGCAAAATTCGGGAACACGCTCGTTGAGTGGTG 678
QY 721 CAGGCGACCAATTTCTCAACTGGGTTCGATCAAAATTCGCGGTAGCTCAAAACGATAGACCT 780
Db 679 AAGGAGA-----TGAGCTCCTACATAAAGAGTCTCGATCCCAACCCCTCGTGGCTGTG 732
QY 781 AACCATATGCTCTCGGTGCGGACGAGGTTCTACATTTGGGTCAACGAGGAGGAAGGAG 840
Db 733 GCGGACGAAGGATTTCTTACGCAACTACCAAGGAATTCAAACCTTACCGTGGAGAAGCCGAG 792
QY 841 TGGCCATACAACGACCGCTCCGACGCGCTCGACACAAATGCTCTTCTCCGTGTCAAGAAC 900
Db 793 TGGGCTTACAGCGGTGGTCT---CGGTGTGACTGGAAGAGCTCTCTTTCGATAGAGAGG 849
QY 901 ATTGACTTTGGCAGCTATCACCTGTACCCGAATTTACTGGGG-----CCAGAAC 948
Db 850 GTGGACTTCGGCAGCTTCCACCTCTATCCGTCCTCCACTGGGTGTTCAGTCCAGAGAATAT 909
QY 949 GCGGACTGGGGAACGCAATGGATCAAGGATCATATTGCCAATGCCCGCAGCGGATCGGGAAG 1008
Db 910 GCCCAGTGGGGAGCAAGTGGATGAAGACCAACATAAAGATCGCAAAAGAGATCGGAAA 969
QY 1009 CGGACCATTCGAGAATTCGG-----CTGGCAGACACCGGACCGGATTCGGTC 1059
Db 970 CCCGTGTTCTGGAAGATATGGAATTCGAAGAGTCCGCCAGTTAACAGACGGGCATC 1029
QY 1060 TATCAGACTGGACCCAGACTCTGGCTAGCAACGGTGAAGCAGGCTTGAACCTTCTGCATG 1119
Db 1030 TACAGACTCTGGAAGCATCTGGTCTACGATCTCGGTGGAGATGGAGCGATGTTCTCGATC 1089
QY 1120 CTCGCTGGGAATGTCACCG 1139
Db 1090 CTCGCGGGAATCGGGAGG 1109

```

```

RESULT 12
LOCUS     AR205088
DEFINITION Sequence 11 from patent US 6368844.
ACCESSION AR205088
VERSION   AR205088.1
KEYWORDS  AR205088.1 GI:21502582
SOURCE    Unknown.
ORGANISM  Unknown.
REFERENCE 1 (bases 1 to 2043)
AUTHORS   Bylina,E.J.
TITLE     Glycosidase enzymes
JOURNAL   Patent: US 6368844-A 11 09-APR-2002;
           Location/Qualifiers
           1..2043
           /organism="unknown"
BASE COUNT 609 a 417 c 592 g 425 t
ORIGIN

```

```

Query Match      5.5%; Score 126.4; DB 6; Length 2043;
Best Local Similarity 49.3%; Prod. No. 6.4e-08;

```


Matches 513; Conservative 0; Mismatches 476; Indels 51; Gaps 5;

QY	121	TTCTGTCACCCATCCCGCGGTGAGTTTCGTCTGAACGGCTTCCCTATCGTTACGGGGGA	180
Db	100	TTCTGTCAAAGTGGAAACGGAATTCGCTCTCAACGGAAGAATTCAGATTCATTGGA	159
QY	181	ACGAACAATATTTACCTGACCTATCAGTCCGACGCCGCTGATGACGCTGTGGCCCAAG	240
Db	160	ACGAACAATATCTATGATGACCTACAGAGCAACGGAATGATAGACAGTGTCTCGAGAGT	219
QY	241	GCTCAAGCCATGAATCTTCTGTCTATCCGACCTGGGTTTTCATCCACATCGSCTCTCT	300
Db	220	GCAGAGACATGGGTATAAAGTCTCTCAATCTGGGTTTCTCCAGCGGAGAGATTAC	279
QY	301	GACGGTCCGTCGCCCAATGATGCGCAACAAGACGGCTTCTACTTTTCAGTACTCGGAC	360
Db	280	TCAGAGACAAGAACACCTACATGCTCTGAGCCCGGTGTTTCGGGGTGCAGAGAAGGA	339
QY	361	CCGTCGACCGCGCTCCGCGCTACAGACGCGGCCACCGCTTGCAGCGCTTGACTAC	420
Db	340	ATATCGAAGCC-----CAGAGCGGTTTCGAAAGACTCGACTAC	378
QY	421	GGGATCCGACGCGCGCGCCGCTTCGGGTCATTGCTCTCTCCTACCAACGACTG	480
Db	379	ACAGTTGCCAAAGAAAGAACTCGGTATAAACTTGTCTATCTGTTGTGAACAACCTGG	438
QY	481	AAAGAAATTTGGGGATGATCAATACGACAAGTGGTACGGCTTCTCTTACCACGACAAC	540
Db	439	GACGACTTCGGTGAATGAACAGTACGTGAGGTGGTGGAGGAACCCATCAGCAGAT	498
QY	541	TTCTACACGACCCCGGACCGAGCGGTACAGAAATGGGTCAATCATCTACTGAAC	600
Db	499	TTCTACAGAGATGAGAAGATCAAGAAGATACAAAGTACGTCTCTTCGTCTAAAC	558
QY	601	CGGTCACACGAGTTACCGCGCTGAGCTACAAGAAGATCCCAACGATCTTGTCTGGGAA	660
Db	559	CATGTCATACCTACAGCGGAGTCTCTTACACGGAAGACCCACCATCATCGCTGGGAG	618
QY	661	CTTGCAATGACCGCGTGGCTAGAGACGGCGACATTAACCACTCGGCGACGTGCAC	720
Db	619	CTTGCAACGAAACCGCGTGTGAGACGGGACAAATCGGGGAACACGCTCTTCAGTGGGTG	678
QY	721	CAGCGCACCATTTGTCACTGGCTCGATCAAAATGTCGGCTACGTCAAAACCATACACCT	780
Db	679	AGGAGA-----TGAGTCTTACATAAAGACTCTGGATCCCAACACCTCTGCGTGTG	732
QY	781	AACCATATGGTCTCGGTGGCGACGAAGGTTTCTACATTGGGTCAACGCGGGAAGCGGC	840
Db	733	GGGACGAAGGATTTCTCAGCAACTACGAAGGATTCARACCTTACGTCGACAAACCGGAG	792
QY	841	TGGCCATACAAGACCGGTCCGACGGGTGACACAACATGCTCTTCTCGGTGTCAAGAAC	900
Db	793	TGGGCTACACGCGGTGCTC---CGGTGTGACTGGAAGAAGCTCTTCGATAGAGCG	849
QY	901	ATTGACTTTGGGACCTATACCTGTACCGGAATTAAGTCTGGG-----CCAGAAC	948
Db	850	GTGACTTTCGGGACGTTCCACCTCTATCCGTCGCCACTGGGTGTGAGTCCAGAGAACTAT	909
QY	949	CGGACTGGGAAACCAATGGATCAAGATCATATTGGCAATGCCGACGATCGGCAAG	1008
Uc	910	CGCCAGTGGGACGAAGTGGATACAAGACCAACATAAAGATCGCAAAAGAGATCGGAAA	969
QY	1003	CCGACCATTCGGAAGATTCGG-----CTGGCAGACACCGGACCGGATTCGCTC	1059
Db	970	CCGCTGTGTTCTGGAGAATATGGAAATCCAAAGAGTGGCGCCAGTTTAAAGAACGCGCATC	1029
QY	1060	TATCAGAGCTGACCCACACTCTCGCTACGAACGCTGAAGCAGGCTGGAACCTCTGGATG	1119
Db	1030	TACAGACTCTGGAAGATCTGGTCTACCATCTCGGTGGAGATGGAGCGATGTTCTGGATG	1089
QY	1120	CTCGCTGGGAATGTCAACGG 1139	
Db	1090	CTCGCGGAATCGGGGAAG 1109	

RESULT 13	AE001779/c
LOCUS	Thermotoga maritima section 91 of 136 of the complete genome.
DEFINITION	Thermotoga maritima section 91 of 136 of the complete genome.
ACCESSION	AE001779
VERSION	AE000512
KEYWORDS	AE001779.1 GI:4981777
SOURCE	Thermotoga maritima.
ORGANISM	Thermotoga maritima.
REFERENCE	Bacteria; Thermotogae; Thermotogales; Thermotogaceae;
AUTHORS	Thermotoga.
	1 (bases 1 to 20990)
	Nelson, K.E., Clayton, R.A., Gilli, S.R., Gwinn, M.L., Dodson, R.J.,
	Haft, D.H., Hickey, E.K., Peterson, J.D., Nelson, W.C., Ketchum, K.A.,
	McDonald, L., Utterback, T.R., Malek, J.A., Linher, K.D., Garrett, M.W.,
	Stewart, A.M., Cotton, M.D., Pratt, M.S., Phillips, C.A.,
	Richardson, D., Heidelberg, J., Sutton, G.G., Fleischmann, R.D.,
	Eisen, J.A., Fraser, C.M. et al.
	Evidence for lateral gene transfer between Archaea and bacteria
	from genome sequence of Thermotoga maritima
	Nature 399 (6734), 323-329 (1999)
	99287316
	-10360571
	2 (bases 1 to 20990)
	Nelson, K.E., Clayton, R.A., Gilli, S.R., Gwinn, M.L., Dodson, R.J.,
	Haft, D.H., Hickey, E.K., Peterson, J.D., Nelson, W.C., Ketchum, K.A.,
	McDonald, L., Utterback, T.R., Malek, J.A., Linher, K.D., Garrett, M.W.,
	Stewart, A.M., Cotton, M.D., Pratt, M.S., Phillips, C.A.,
	Richardson, D., Heidelberg, J., Sutton, G.G., Fleischmann, R.D.,
	White, O., Salzberg, S.L., Smith, H.O., Venter, J.C. and Fraser, C.M.
	Direct Submission
	Submitted (01-JUN-1999) The Institute for Genomic Research, 9712
	Medical Center Dr, Rockville, MD 20850, USA
FEATURES	Location/Qualifiers
source	1. .20990
gene	/organism="Thermotoga maritima"
	/db_xref="taxon:2336"
	/complement(97..1203)
CDS	/gene="TM1224"
	/complement(97..1203)
	/gene="TM1224"
	/note="similar to PID:1208894 SP:Q44406 percent identity:
	54.86; identified by sequence similarity; putative"
	/codon_start=1
	/transl_table=1
	/product="transcriptional regulator, XylR-related"
	/protein_id="AAD36299.1"
	/db_xref="GI:4981778"
	/translation="MPKSVRAENISRIILKRIKSPVSRVELAEELGLTKTTVEIAKI
	FLEKGLVVEKDSKGVGRPTKSLKISPCAYVLGIEVTRDEIAACLDASNNLAHE
	AHPLDSQDRETLNVMRTIDRAKMMKLGKLSALTVAAPGIDTFRGLIIDPRN
	PPLSOLPLANLLKEKYIGIEVVDNDADGCEKWTTRKDSFAWILTKGIGAGII
	DEGLYRGENGAGEIGYTRVFNENGVLEEDYCNENVLKHLVSHMGFSLAARQSGD
	HSVKEYFDIARYFISGLNLIHLFGISKIVIGGFFKELGENFLKKIKIEVTHLLYK
	HSDMSFSKVOEPVIAFGAAVAHALENLYERTVS"
	/complement(1217..2197)
	/gene="TM1225"
	/complement(1217..2197)
	/gene="TM1225"
	/note="similar to GB:Pyro_h percent identity: 52.69;
	identified by sequence similarity; putative"
	/codon_start=1
	/transl_table=1
	/product="conserved hypothetical protein"
	/protein_id="AAD36300.1"
	/db_xref="GI:4981779"
	/translation="MKVTEKIPNIPEWERPEGYTPVMRYSKNPICIGNPVPKCARV
	FNSAVPVNGEFGVFRIDHKNTBPLHFGSKDGINWIEPEEIQWVDVNGEPQPS
	YADPRVKIEDTYITFTCTDDIGFTICGWTIKDFKTEFLPNTVPYRNGVLPFRK
	INGKYVMLNRDNGHTPFQDIFLSESDPMTHWGNHREVLGRSSYNWENKICAGPY
	PIETSEGWLLIYHGVTLTCNGYVYFSGAALLDLDPSKVLRSRYLLTPEEVEVTG

Db	5758	GCAGAGACATGGGTATATAAGGTCTCAGAACTCGGGTTTCTCGACGGGAGAGATTAC	5699
Qy	301	GACGGCTCGTCCCAACATTCGATGCGCAACAACAGCGCTTCTACATTCTCAGTACTCGGAC	360
Db	5698	TCCAGAGACAAGAACACCTACATCGATCTCGACCGCGGTCTTTCCGGGTGCCAAGAGGA	5639
Qy	361	CGGTCCAGCGGGCTCCGGCGGTACAAACGACGGCGCCACCGGCTTGAAGCCCTTGACTAC	420
Db	5638	ATATCGAACGCC-----CAGACGGTTTTCGAAAGACTCGACTAC	5600
Qy	421	CGCATCGCAGCGCGCGCGCGCGCTTCGGGTCATTTCTGCTCTCCTCACCACGACTGG	480
Db	5599	ACAGTTGCGAAGCGGAAGAACTCGGTATATAAACTTGTCTATGTTCTTGTGAACAACCTGG	5540
Qy	481	AAAGAAATTTGGGGAATGATCAATACGACAAGTCTGACGGCTTCTCTTACCACGACAAC	540
Db	5539	GACGACTTCGGTGAATGAACAGTACGTGAGGTGTTGGAGCAACCCATCACGAGAT	5480
Qy	541	TTCTACACGACCCCGGACCCAGCAGCGGTACAAGATTGGGTCAATCATCTACTGAAC	600
Db	5479	TTCTACAGAGATGAGAAGATCAAGAAGAGTACAAAAGTACGTCTCTTCTCGTAAAC	5420
Qy	601	CGGTCAACAGCATTTACCGCGGTGACGTACAAGACGATCCCAACGATCTTTGCTGGGAA	660
Db	5419	CATGTCATACCTACACGGGAGTTCCTTACAGGGAAGAGCCCAACCATCATGCGCTGGAG	5360
Qy	661	CTTGCAATGAGCGCGCTCCGTAGGAAGCGGCACATTACCAACCTCGGCGACGTGCAC	720
Db	5359	CTTGCAAAACGACCGCGCTGTGAGACGGACAATTCGGGGAACACGCTCGTTGAGTGGTG	5300
Qy	721	CAGGCGACCATTTCTCACTGGGTGATCAAAATGTCGGCGTACGTCAAAAGCATAGACCT	780
Db	5299	AAGGAGA-----TGAGCTCTACATAAAGAGTCTCGATCCCAACCACTCTGTGGCTGTG	5246
Qy	781	AACCATATGCTCTCGTCCGCGACGAAGGTTCTACATTGGGTCAACGAGGAAACCGCG	840
Db	5245	GGGGACGAAGGATTTCTTACGCAACTACGAAGGATTTCAAACCTTACGGTGGAGAAGCCGAG	5186
Qy	841	TGGCCATACAAGACCCGCTCCGACCGGTGCGACAACATGCTCTTCTCGGTGTCGAAGAC	900
Db	5185	TGGCGCTACAACCGCTGGTC---CGGTGTTGACTCGGAAGACTCTCTTCGATAGAGACG	5129
Qy	901	ATTGACTTTGGCAGCTATCACCTGTACCCGATTAATCTACTGGG-----CCAGAAC	948
Db	5128	GTGGACTTCGGCAGCTTCACCTCTATCGCTGCCACTGGCGGTCTCAGTCCAGAGAATAT	5069
Qy	949	CGCGACTGGGGAACGCAATTCAGATCATATTGCGAATGCGCGCAGCGATCGGCAAG	1008
Db	5068	GCCAGTGGGAGCGAAGTGGATAGAGACCACATAAAGATCGAAGAGAGATCGGAAA	5009
Qy	1009	CGACCATTTCTCGAAGATTCGG-----CTGGCAGACACCGGACCGGATTCGGTC	1059
Db	5008	CCCGTTGTTCTGGAAGATATGGAATTCCAAAGAGTGCGCCAGTTAACAGAACGGCCATC	4949
Qy	1060	TATCAGACTGACACCACTGCTGCTAGCAACGGTGAAGAGGCTGGAACCTTCTGGATG	1119
Db	4948	TACAGACTCTGGAAGCATCTGCTCTACGATCTCGGTGGAGATGGAGCGATGTTCTGGATG	4889
Qy	1120	CTCGCTGGGAATGTCAACGG	1139
Db	4838	CTCGCGGGAATCGGGAAGG	4869
RESULT	14		
AY033477			
LOCUS		2001 bp	DNA linear
DEFINITION	AY033477	Thermotoga neapolitana	
ACCESSION	AY033477	beta-mannanase (man5) gene, complete cds.	BCT 09-FEB-2002
VERSION	AY033477.1	GI:18642981	
KEYWORDS			
SOURCE		Thermotoga neapolitana.	
ORGANISM		Thermotoga neapolitana	
		Bacteria; Thermotogae; Thermotogales; Thermotogaceae;	
REFERENCE	1 (bases 1 to 2001)		
AUTHORS	Parker, K.N., Chhabra, S.R., Lam, D., Callen, W., Duffaud, G.D., Sneed, M.A., Short, J.M., Mathur, E.J. and Kelly, R.M.		
TITLE	Galactomannanases Man2 and Man5 from Thermotoga species: growth physiology on galactomannans, gene sequence analysis, and biochemical properties of recombinant enzymes		
JOURNAL	Biotechnol. Bioeng. 75 (3), 322-333 (2001)		
MEDLINE	21473892		
PUBMED	11590605		
REFERENCE	2 (bases 1 to 2001)		
AUTHORS	Parker, K.N., Chhabra, S.R., Lam, D., Callen, W., Duffaud, G.D., Sneed, M.A., Short, J.M., Mathur, E.J. and Kelly, R.M.		
TITLE	Direct Submission		
JOURNAL	Submitted (26-APR-2001) Chemical Engineering, North Carolina State University, Room 3309, 840 Main Campus Drive, Raleigh, NC 27606, USA		
FEATURES	Location/Qualifiers		
source	1..2001		
	/organism="Thermotoga neapolitana"		
	/strain="DSM 5068"		
	/db_xref="taxon:2337"		
gene	1..2001		
	/gene="man5"		
CDS	1..2001		
	/gene="man5"		
	/codon_start=1		
	/transl_table=1		
	/product="beta-mannanase"		
	/protein_id="AAK53459.1"		
	/db_xref="GI:18642982"		
	/translation="MKLYPSFLIVTLPIVLFANSDFVKVENCGRFLUNGEEFRFVGSN NTYMHYKSRMIDSVLESAMGVKVLRIWGLDGBESYCDKNTYMHAPVFGFLPEG TAAQGFERLDYTVAKAKELGLIILVNLNDDFGMNOYVRMFGGIHHDDFYRNEK IKKEYKKYVFLINVTYTPVREPTIMAEELANPRCETDKSGNTLVEWVEKMS AYIKSLDPNHLVAVGDEGFNNYEGRPYGAEMAYNGMVGDKRLLEIEIVDFCT FHLTPSHGCVSPENVAOMCAKIEDHIKIAKEVGKPVLEEYGIKPSAPNVAIYKL WDLVYNLGNGCAFMWLAGIGEGDRDEKGYPDYDGFRIYNDESEAKLIREYAKL FSTGDTREDTCTFTPKDQGEIKTKTVKVRGVDFYSNTFKGISVGVNLLIFDEIKH LGYVYGFEDTTRISDGEHMFLEAHFGETVKDITRVKVNRAQYVLAHEDFSRP EVKMNNSGTWOAEFTKPTDIEMNGEVNGALOMVYLPCKGDMESVRVVRKKDLPV CSILLEYDIYIPDVEGLTRLRPYAVLNPGLWIKGLDMNNTSIDSGELVFDGKKYKF HVRIEFDKTPGVNELHIGVVDHLEVDGFIIDNRLYKKS"		
BASE COUNT	598 a	394 c	580 g 429 t
ORIGIN			
Query Match	5.4%	Score 124.4;	DB 1; Length 2001;
Best Local Similarity	53.0%	Prod. No. 1.2e-07;	
Matches	424;	Conservative	0; Mismatches 316; Indels 60; Gaps 5;
Qy	376	CCGGCGCTACACGACCGCGCGCGCTTCCAGGCGCTTCAAGCGCTTCACTACGCGATCGCGAGCGCG	435
Db	298	CCAGAGGGTACGAACTCAGGACGGTTTGAAGAGCTCCAGCTACGAGTAGCGAAGCA	357
Qy	436	GCGCGCCAGCGCTTCGGGTGATTGTCCTCCACCAACGACTGGAAGAATTTGGGGA	495
Db	358	AAAGAACTCGGCATAAAGCTCATAATCGTTCTTGTGAACAACTGGGACGACTTCGGGTGGA	417
Qy	496	ATGGATCAATACGACAAGTGGTACGCGCTTCTTACCACGACAACCTTCTACACGCGACCC	555
Db	418	ATGAATCAATACGTGAGATGGTTTGGGGGCATCCATCAGCATCTCTACAGGAACGAG	477
Qy	556	CGGACCCACGCGCTACAGAATTCGGTCAATCATCTACTGAACGGGTCAACAGCATTT	615
Db	478	AAAGTAAAGAAGATACAAAAGTAGTCTCTTCTCTCTATAAACAGGTTGAACACCTAC	537
Qy	616	ACCGCGCTGACGTACAGAAGCATCCAAAGTCTTCTTACCACGACAACCTTCTACACGCGCC	675
Db	538	ACGGGTGTTCTTACAGGAAGAGCCACCATCATCGCATGGCACTGGCAGACGAGCCC	597
Qy	676	CGCTGCGTAGGAAGCGGCACATTTACCAACCTCGGGCACGTGCACTCAGGCGACGATTTGC	735
Db	598	AGGTGTGAACGCGAC-----AAGTCTGGTAAACACACTCGTT	633

QY 736 AACTGGGTTCGATCAATATGCGCGTACGTCACAAAGCATAGACCCCTAACCATATGCTCTCG 795
DB 634 GAATGGGTAGAGAGATGAGTCTTACATAAAGAGTCTGTGATCCAAACCCACCTGGTTC 693
QY 796 GTCCGGGACGAGGGTTCTACA-----TTCCGCTCAACGCGCAGGAGCGGCT 841
DB 694 GTGGGAGACGAGGATTTCTTCAACAACACTACGAGGCTTCAGACCTTACCGTGGAGAGCT 753
QY 842 G-CCCATACAACGACCCGTCACCGGCTCGACACAATGCTCTCTCTCGTCTCAAGAAC 900
DB 754 GAGTGGGCTACAACGGATGGTCCCGTGTGTGAGTGAAGAGACTTCTGAGATAGAGCG 813
QY 901 ATTGACTTTGGCAGCTATCACCTGTACCCGAATTAAGTGGGGCAGC-----AC 948
DB 814 GTGGATTTTGGTACGTTCCATCTCTACCCCTCCCACTGGGGTGTGAGCCCTGAAGAACTAC 873
QY 949 GCGGACTGGGGAACCAATGGATCAAGGATCATATTTGCAATGCGCAGCGATCGGCAAG 1008
DB 874 GCACAGTGGGGGCAAGTGGATCAAGATACATAAAGATCCCAAAAGAGGTGGAAAA 933
QY 1009 CGACCATTTCTCGAAGATTCGGCTGGCAGACACCG-----ACCGCGATTCCGTC 1059
DB 934 CCGCTGCTTCTGAGAGACTACGGTATTCCTCAAAAGTCCCGGTCACACAGGGTTGCCATT 993
QY 1060 TATCAGACGTGGACCCAGACTGTGCGTACGAACGCTGAAGCAGGCTGGAATCTCTGATG 1119
DB 994 TACAAATTGTGGAACCATCTGGTCTACACCTCGGTGGAAACGGTGCCATGTTCTCGATG 1053
QY 1120 CTCGCTGGCAATGTCAACGG 1139
DB 1054 CTCGCGAGGAATCGTGAAG 1073

RESULT 15
LOCUS TNE17981
DEFINITION Thermotoga neopolitana manB gene.
ACCESSION Y17981
VERSION Y17981.1 GI:6006596
KEYWORDS beta-mannosidase; manB gene.
SOURCE Thermotoga neopolitana.
ORGANISM Bacteria; Thermotogae; Thermotogales; Thermotogaceae;
Thermotoga.
REFERENCE 1 (bases 1 to 2001)
AUTHORS Parker,K.N., Lam,D., Duffaud,G., Snead,M.A., Mathur,E.J. and Kelly,R.M.
TITLE Amino Acid Sequence of beta-mannosidase genes from the hyperthermophilic bacteria Thermotoga maritima and Thermotoga neopolitana
JOURNAL 2 (bases 1 to 2001)
AUTHORS Parker,K.N.
TITLE Direct Submission
JOURNAL Submitted (21-AUG-1998) K.N. Parker, North Carolina State University, Department of Chemical Eng, Box 7905 NCSU, Raleigh, NC 27695-7905, USA
COMMENT sequence incorrect.
FEATURES
source
1..2001
Location/Qualifiers
/organism="Thermotoga neopolitana"
/strain="5068"
/db_xref="taxon:2337"
1..2001
/gene="manB"
1..2001
/gene="manB"
/codon_start=1
/translation="1"
/product="beta-mannosidase"
/protein_id="CAB56856.1"
/db_xref="GI:6006597"

/db_xref="SPTREMBL:O9RIK7"
translation="MKLVFSLPIVLTLPILVLFANSDFKVKNGRFLLNGEELKFGVGSN
NYMHYKSNRMIDSVLESKANGVKVLRWGLDGESECRDKNYTHIPAPVGLPEG
TNAQDGFRLDYVAKAKELGKLIIVLNNHDDFGGMNOYVRWEGGLIUDDFYRNEK
IKEEKYKVSFLINRNTYTGPVREEPTIMAHELANPRCETDKSNTLVNWEVKS
AYIKSLDPNHLVAVGDEGFNNYEGFRPYGGAEAWAYNMGSDVMKRLLEITVDGFT
PHLYPSHMGVSPENYAOMGAKWIEDHIIKIAKEGVKPVLEEGTPKSAVNRVAIYKL
WNDLVYNLGNGAMFWMLAGICEGWRDEKGYYPDYDFRIVNDSEEAALIREYAKL
FSTGDTREDTCMFITPKDGOEIKKTVKVRGVDFYSNTFKGISVGVENLLPEDEIKH
LGYGYGFEEDTTRISDGEHEMFLEAHFRGETVKDTIRVKVNRAGVYLAESVDFSRP
BEVKNNWSGTWQAEFKTPDIEWNGEVGNALOMNVYLPKGDNEEVRVVRKFDQLPV
CEILEYDIYIPDVEGTRIPYAVLNPQWVKIGLDMNNTSIDSGELVSHNKKKYHKF
HYRIEFDKTPGVNELHIGVGDHLEYDGFIFIDNVRLYKKS"

BASE COUNT 598 a 394 c 580 g 429 t

Query Match 5.4%; Score 124.4; DB 1; Length 2001:
Best Local Similarity 53.0%; Pred. No. 1.2e-07;
Matches 424; Conservative 0; Mismatches 316; Indels 60; Gaps 5;

QY 376 CGCGCGTACAACGACGCGCGCGCTTCGAAGGCTTCACTACGCGTGGCAGCGCG 435
DB 298 CCAGAGGTACGAACGCTCAGGACGGTTTGAAGACTCGACTACCGGTAGCGAAGCA 357
QY 436 GCGCGCACGCGCTTCGGGTGATTGTCTCTCCACCACCAACGACTGGAAAGAAATTTGGCGGA 495
DB 358 AAAGAAGCTGGGCATAAAGCTCATAATCGTTCTTGTGAACAACCTGGGACGACTTCGGTGA 417
QY 496 ATGGATCAATACGACAAGTGGTACGCGCTTCCTTACCACACAACCTTCTACACCGACCCC 555
DB 418 ATGAATCAATACGTGAGATGGTTTGGGGGATCCATCAGCATCTCTACAGGAAGCAG 477
QY 556 CGGACCCAGCGGTACAAGAAATTTGGTCAATCATCTAGTGAACGGGTCAACAGCATTT 615
DB 478 AAGTCAAGAAGAAATACAAAAGTACGTCTCTCTCATAAACAGGTGAACACCTAC 537
QY 616 ACAGCGGTACGTACGAAGACGATCAACGATCTTGTCTTGGGAACCTTGGCAATGAGCGG 675
DB 538 ACAGGTCTCTTACAGGGAAGAGCGCACCATCATGCGATGGGAACCTGGGGAAGAGTCC 597
QY 676 CGCTCGGTAGGAGCGGCACATTTACCANCTCGGCACCTGCACTACGGGCAATCTATTC 735
DB 598 AGGTGTGAACAGGAC-----AAGTCTGTGAACACACTCGTT 631
QY 736 AACTGGGTTCGATCAATATGCGCGTACGTCACAAAGCATAGACCCCTAACCATATGCTCTCG 795
DB 634 GAATGGGTAGAGAGATGAGTCTTACATAAAGAGTCTGTGATCCAAACCCACCTGGTTC 693
QY 796 GTCCGGGACGAGGGTTCTACA-----TTGGGTCAACGCGCAGGAGCGGCT 841
DB 694 GTGGGAGACGAGGATTTCTTCAACAACACTACGAGGCTTCAGACCTTACGTTGGAGAGCT 753
QY 842 G-CCCATACAACGACCCGTCGCGGCGTGGACACAATGCTCTCTCTCGTCTCAAGAAC 900
DB 754 GAGTGGGCTACAACGGATGGTCCCGTGTGTGAGTGAAGAGACTTCTGGAGATAGAGCG 813
QY 901 ATTGACTTTGGCAGCTATCACCTGTACCCGAATTAAGTGGGGCAGC-----AC 948
DB 814 GTGGATTTTGGTACGTTCCATCTCTACCCCTCCCACTGGGGTGTGAGCCCTGAAGAACTAC 873
QY 949 GCGGACTGGGGAACCAATGGATCAAGGATCATATTTGCGGAATCCCGGACCGATCGGCAAG 1008
DB 874 GCACAGTGGGGGCAAGTGGATAGAGATACATAAAGATCCCAAAAGAGGTGGAAAA 933
QY 1009 CGACCATTTCTCGAAGAAATTCGGCTGGCAGACACCG-----ACCGCGATTCCGTC 1059
DB 934 CCGCTGCTTCTGGAAGAGTACGGTATTTCCAAAAGTCCCGGTCACACGGTTGCCATT 993
QY 1060 TATCAGACGTGGACCCAGACTGTGCGTACGAACGCTGAAGCAGGCTGGAATCTCTGATG 1119
DB 994 TACAAATTGTGGAACCATCTGGTCTACACCTCGGTGGAAACGGTGCCATGTTCTCGATG 1053
QY 1120 CTCGCTGGGAATGTCAACGG 1139

||||| ||| | | ||
Db 1054 CTCGAGGAATCGGTGAAG 1073

Search completed: November 13, 2002, 10:08:04
Job time : 4057 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 13, 2002, 08:58:51 ; Search time 326 Seconds

(without alignments)

15812.344 Million cell updates/sec

Title: US-09-917-378-2

Perfect score: 2289

Sequence: 1 atgggtctagtgcgtcgcc.....tcacctgcacgctagtgtga 2289

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 1125999159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N_Geneseq_101002.*

- 1: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT.*
- 2: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT.*
- 3: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT.*
- 4: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT.*
- 5: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT.*
- 6: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT.*
- 7: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT.*
- 8: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT.*
- 9: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT.*
- 10: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT.*
- 11: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT.*
- 12: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT.*
- 13: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT.*
- 14: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT.*
- 15: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT.*
- 16: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT.*
- 17: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT.*
- 18: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT.*
- 19: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT.*
- 20: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT.*
- 21: /SID52/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT.*
- 22: /SID52/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT.*
- 23: /SID52/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT.*
- 24: /SID52/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	180.4	7.9	3004	17	AAT12337
2	180.4	7.9	3004	21	AAZ55924
3	180.4	7.9	3004	24	ABK86729
4	143.6	5.8	11706	20	AAZ55661
5	143.6	5.8	11707	24	AAD28525
6	128.4	5.6	6415	20	AAZ55662
7	128.4	5.6	6416	24	AAD26526
8	128	5.6	2043	18	AAT93688
9	128.4	5.5	1992	19	AAV36921

10	126.4	5.5	2043	19	AAT94213	Thermotoga maritim
11	124.8	5.5	2043	19	AAV36917	Thermotoga maritim
12	124.4	5.4	2001	19	AAT94200	Thermotoga neapolit
13	114.8	5.0	2977	12	AAO13001	Endol gene encodin
14	94.6	4.1	1621	19	AAV07163	Thermomonospora fu
15	90.2	3.9	2029	20	AAZ55660	DNA sequence encod
16	90.2	3.9	2029	24	AAD26568	Active cellulase h
17	89	3.9	2600	12	AAO15178	Portion of PAEC-1
18	82	3.6	777	21	AAC44444	Arabidopsis thalia
19	82	3.6	784	21	AAC47999	Arabidopsis thalia
20	82	3.6	796	21	AAC45005	Arabidopsis thalia
21	82	3.6	2011	21	AAC52204	Arabidopsis thalia
22	81.6	3.6	4056	22	AAH75060	Nucleotide sequenc
23	78.4	3.4	390	13	AAO21833	Randomising oligon
24	78.4	3.4	390	14	AAQ36859	PCR primer for 5'
25	78.4	3.4	390	22	AAQ36859	Sequence containin
26	78.4	3.4	390	22	AAI72775	Oligo #7 for clon
27	77	3.4	4059	22	AAH75059	Nucleotide sequenc
28	75.8	3.3	2293	19	AAV09659	A. cellulolyticus
29	72.6	3.2	38186	20	AAZ32028	Human KTH1 relate
30	72.6	3.2	38186	22	AAQ90085	AC004449 cDNA clon
31	70.6	3.1	2308	24	ABQ39168	Oligonucleotide fo
32	70.6	3.1	2308	24	ABQ39169	Oligonucleotide fo
33	70.6	3.1	6510	22	AAZ46648	Tumour suppressor
34	70.6	3.1	6510	24	ABL32395	Human immune syste
35	70.6	3.1	6510	24	AAZ61115	Human gene regulat
36	69.8	3.0	1232	22	AAI11124	Human small cell l
37	69.6	3.0	2175	14	AAO49820	NK-1 cellulase gen
38	69.6	3.0	2365	24	ABQ17066	Oligonucleotide fo
39	69.6	3.0	2365	24	ABQ17067	Oligonucleotide fo
40	69.2	3.0	150	20	AAV64956	Mouse histone H2B
41	68.8	3.0	1314	24	ABK73393	Bacillus lichenif
42	68.2	3.0	1844	21	AAC41572	Arabidopsis thalia
43	68.2	3.0	1844	21	AAC47098	Arabidopsis thalia
44	68.2	3.0	1847	21	AAC38734	Arabidopsis thalia
45	65.6	2.9	2510	24	ABK53202	Bacillus subtilis

ALIGNMENTS

RESULT 1
AAT12337
ID AAT12337 standard; DNA: 3004 BP.

XX AAT12337;

XX 08-OCT-1996 (first entry)

XX A. cellulolyticus E1 endoglucanase gene.

XX E1 endoglucanase; cellulase; cellulose; saccharification; ethanol:

XX SS.

XX Acidothermus cellulolyticus.

XX Key Location/Qualifiers

FT misc_difference 200

FT /note= a

FT /tag= b

FT misc_signal 710..725

FT /note= putative upstream regulatory sequence"

FT RBS 772..779

FT /tag= c

FT repeat_unit 781..790

FT /tag= d

FT /rpt_type= DIRECT

FT repeat_unit 795..804

FT /tag= e

FT /rpt_type= DIRECT

base g at position 200 is shown in Fig 1, but is missing in the sequence given on page 22 of the specification"

```

FT repeat_unit 810..817
FT /*tag= f
FT /rpt_type= DIRECT
FT CDS 824..2512
FT /*tag= g
FT sig_peptide 824..946
FT /*tag= h
FT sig_peptide 863..946
FT /*tag= i
FT /*note= "alternative signal sequence"
FT mat_peptide 947..2509
FT /*tag= j
FT terminator 2514..2560
FT /*tag= k
FT PN W09602551-A1.
FT XX
FT PD
FT PO
FT PF 14-JUL-1995; 95WO-US08868.
FT XX
FT PR 15-JUL-1994; 94US-0276213.
FT PR 26-SEP-1989; 89US-0412434.
FT PR 27-JAN-1992; 92US-0826089.
FT PR 21-SEP-1993; 93US-0125115.
FT XX
FT PA (MIDE ) MIDWEST RES INST.
FT XX
FT PI Himmel ME, Laymon RA, Thomas SR;
FT XX
FT DR WPI; 1996-105843/11.
FT DR P-PSDB; AAR89927.
FT XX
FT PT New isolated DNA encoding endoglucanase - obtd from Acidothermus
FT cellulolyticus, used for prodn of the enzyme for use in cellulose
FT hydrolysis.
FT XX
FT PS Claim 4; Page 22-24; 34pp; English.
FT XX
FT CC The E1 endoglucanase gene (T12337) of Acidothermus cellulolyticus
FT codes for an enzyme, the mature form of which (R89927) can hydrolyse
FT cellulose and xylan. The gene was obtd. by ligating Sau3A fragments
FT of A. cellulolyticus genomic DNA into EMBL3 lambda phage arms,
FT CC packaging the phage DNA, plating on E.coli LE392 agar contg. CMC,
FT staining with Congo Red, and isolating DNA from positive clones.
FT CC Alternative signal sequences were identified (see also R89928-29).
FT CC The gene can be incorporated into a vector and used for large-scale,
FT low-cost prodn. of recombinant E1 endoglucanase, pref. using
FT CC Saccharomyces, Zymomonas or E.coli hosts. The enzyme is useful
FT for the saccharification of cellulosic biomass for fermentation to
FT ethanol.
FT XX
FT SQ Sequence 3004 BP: 604 A; 931 C; 938 G; 531 T; 0 other;

Query Match 7.9%; Score 180.4; DB 17; Length 3004;
Best Local Similarity 64.8%; Pred. No. 3e-25;
Matches 284; Conservative 0; Mismatches 151; Indels 3; Gaps 1;

QY 1852 CCAACCCGTCGCCGAGCCGCGAGCCCAAGCCCGAGTCCTCCCATCCCGTCC 1911
DB || ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 2078 CCGAGCCGTCGCGAGTCGGAGCGGAGCGCTACTCCGCGCGACGCGCGAGC 2137
DB || ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1912 CCGAGCCGCGAGCCGAGCCGTCGCGTCGCCGCGCGAGTCGCGCGAGT 1971
DB || ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 2138 CCAACGCTGACCCCTACTGCTACGCGCCGCGCGAGCCGCGAGCCGCGAGC 2195
QY 1972 CTGTCGTCGTCGGGTGTCGGGCGCGAGCTATGTGGTCAATAGTATTGGGTTCT 2031
DB || ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 2196 -CGGAGCGCTCGGAGCCGCTGCACCGGAGTTACCGAGTCAACGAGATTGGGCAAT 2254
DB || ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 2032 GCGTTTACGGGACCGGTGACGCTGACGATACCGGAGCCGCGAGCGGGTGGAGC 2091
DB || ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 2255 GCGTTTACGGGTAAACGGTGCCGCTGACAAATTCGGATCCGTCGCGACCAAGACATGGAGC 2314
DB || ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

```

QY 2092 GTGCGGTGTCGTTTGGTGGGAATCAGACGCTCAGCAACTACTGGAACACTGCGTTGACC 2151
DB || ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 2315 GTCAGTTGGACATTCGCGGAAATCAGACGATTAACCAATTCTGGAATCGAGCGGTACG 2374
DB || ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 2152 CAATCAGGTGCATCGGTGACGCGGAGCACTGAGCTACAAACAGTGATCCCAACCGGT 2211
DB || ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 2375 CAGAAGCGTCAGTCGGTAAAGCGCTCGGATATGAGTTATAACAACGTTGATTGAGCCTGGT 2434
QY 2212 CAGTCGACCACTTCGGATTCAACGGGAGTTACTCAGGAACAACACCGCACCTACACTC 2271
DB || ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 2435 CAGAACCAACGTTTCGAGATTCAGGCGAGCTATACCGGAAGCAACGCGGACCGACAGTC 2494
QY 2272 ACCTGCACGCTAGTTGA 2289
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 2495 GCCTGCGCAGCAAGTTAA 2512
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 2
AAZ55924
ID AAZ55924 standard; DNA; 3004 BP.
XX
AC AAZ55924;
XX
DT 10-APR-2000 (first entry)
XX
DE Acidothermus cellulolyticus E1 endoglucanase gene.
XX
KW E1 endoglucanase; cellulose binding domain; CBD; cellulose modification;
KW beta-1,4-endoglucanase; endocellulase; thermostable; ds.
XX
OS Acidothermus cellulolyticus.
XX
FH Key Location/Qualifiers
FT CDS 824..2512
FT /*tag= a
FT sig_peptide 824..946
FT /*tag= b
FT sig_peptide 863..946
FT /*note= "Encodes putative signal peptide"
FT mat_peptide 947..2509
FT /*tag= c
FT /*note= "Encodes putative signal peptide (alternative)"
FT misc_feature 2204..2509
FT /*product= "Mature E1 endoglucanase"
FT /*tag= f
FT /*note= "Encodes cellulose binding domain (specifically
FT claimed)."
XX
PN CA2226898-A1.
XX
PD 25-SEP-1999.
XX
PF 25-MAR-1998; 98CA-2226898.
XX
PR 25-MAR-1998; 98CA-2226898.
XX
PA (MIDE ) MIDWEST RES INST.
XX
PI Laymon RA, Adney WS, Thomas SR, Himmel ME;
XX
DR WPI; 2000-087663/08.
DR P-PSDB; AAY69508.
XX
PT Isolated domains of Acidothermus cellulolyticus E1 endoglucanase useful
PT for labeling or modifying a cellulose and for purifying or immobilizing
PT a binding domain fusion protein to cellulose.
XX
PS Claim 12; Fig 1; 85pp; English.
XX
CC This sequence represents the Acidothermus cellulolyticus E1 endoglucanase
CC gene, encoding a beta-1,4-endoglucanase, or endocellulase. The cellulose

```


binding domain (CBD) of EI endoglucanase, and nucleotides which encode it are specifically claimed. The CBD is believed to be roughly wedge-shaped; as the CBD binds to cellulose, it is thought that the wedge tip is inserted between the microfibrils of the cellulose fibre, disrupting the crystalline structure, and making the cellulose linkages more accessible to the catalytic domain of the EI endoglucanase. The EI endoglucanase CBD is useful in labelling or modifying the surface of cellulose or other polysaccharides. Such modified cellulose can then be used in textile, pulp, paper, chemical and pharmaceutical industries. CBDs can be used in affinity purification of CBD-fusion proteins, and can also be used to immobilise the CBD-fusion proteins to a cellulose support. CBD-fusion proteins can be used to modify the chemical or physical properties of a cellulose or polysaccharide matrix column and to modify (e.g., roughen or disrupt) a cellulose or polysaccharide fibre. The CBD of the EI endoglucanase exhibits greater stability at pH 4-8 and has an optimum temperature for stability of 83 degrees Celsius which is not found in CBDs from non-thermophilic organisms.

Sequence 3004 BP; 605 A; 933 C; 935 G; 531 T; 0 other;

Query Match 7.9%; Score 180.4; DB 21; Length 3004;
Best Local Similarity 64.8%; Pred. No. 3e-25;
Matches 284; Conservative 0; Mismatches 151; Indels 3; Gaps 1

QY	1952	CCAACCCCGTCCCGAGAGCCCGGACCCAGCCAGCCAGCTCTCCCATCCCGTGC	1911
Db	2078	CCGAGCCGTCGGGAGCATCGGACGCCGACGCTACTCGAGCCGACAGCCAGCCGACG	2137
QY	1912	CCGAGCCGACGCCAGCCCTACGCCGTCCCGTCGCCAGGCCGTCGCCGTCCCGGAGT	1971
Db	2138	CCAAGCTGACCCCTACTGCTACGCCACGCCACGGCAGCCGCGCTACCCGA--	2195
QY	1972	GTGTCGTGTCGGGTGCGGTGCCGGCGACGTATGTGTGAATAGTCAATTGGGGTTCT	2031
Db	2196	-CGCGAGCCTCCGAGGCCGCTGCACCGGATTACAGGTCAACAGCGATTGGGCAAT	2254
QY	2032	GGGTTTACGGCGAGCGTCACGGTCAGCAATACCGGGAGCCGGGCGACAGCGGGTGGAGC	2091
Db	2255	GGCTTCACGGTAACGGTGGCGCTGACAAATTCGGATCCGTCGCCGCAAGACATGGAGC	2314
QY	2092	GTGCGCTGCTGTTGGTGGGAATCAGACGGTCACGAATCTCTGGAACACTCGCGTTGACC	2151
Db	2315	GTCAAGTTGGACATTCGCGGGAATCAGACGATTACCAATTCTGGTAATGCACGGTCAAG	2374
QY	2152	CAATCAGGTGCATCGGTGACGGCGAGCAACCTGAGCTACAACAGGTGATCCAAACCGGT	2211
Db	2375	CAGAACGGTCAGTCGGTAACGGCTCGGAATATGAGTTATAACACAGTGATTCAAGCTGCT	2434
QY	2212	CAGTCGACACACTTTCGGATTCAACGGAAATTACTCAGGAACAAACACCGCACTACACTC	2271
Db	2435	CAGAACACAGGTTCCGATTCCAGCGGAGCTATACCCGGAAGCAACCGCGGCACGATC	2494
QY	2272	ACCTGCACGGCTAGTTGA	2289
Db	2495	GCCTGCGCAGCAAGTTAA	2512

RESULTS

ABK86729

ID ABK86729 standard; cDNA: 3004 bp.

XX

AC ABK86729;

XX

DE A. cellulolyticus cellulase EI beta-1,4-endoglucanase precursor cDNA, el.
xx 24-SEP-2002 (first entry)

Gene: ss; transgenic: plant; lignocellulose: cellulase; ligninase:

KW fermentable sugar; ethanol; fermentation; silage: feed: fuel:

KW industrial chemical; biodegradation; chloroaromatic;

KW environmental pollutant; EI beta-1,4-endoglucanase; el.

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100	101	102	103	104	105	106	107	108	109	110	111	112	113	114	115	116	117	118	119	120	121	122	123	124	125	126	127	128	129	130	131	132	133	134	135	136	137	138	139	140	141	142	143	144	145	146	147	148	149	150	151	152	153	154	155	156	157	158	159	160	161	162	163	164	165	166	167	168	169	170	171	172	173	174	175	176	177	178	179	180	181	182	183	184	185	186	187	188	189	190	191	192	193	194	195	196	197	198	199	200	201	202	203	204	205	206	207	208	209	210	211	212	213	214	215	216	217	218	219	220	221	222	223	224	225	226	227	228	229	230	231	232	233	234	235	236	237	238	239	240	241	242	243	244	245	246	247	248	249	250	251	252	253	254	255	256	257	258	259	260	261	262	263	264	265	266	267	268	269	270	271	272	273	274	275	276	277	278	279	280	281	282	283	284	285	286	287	288	289	290	291	292	293	294	295	296	297	298	299	300	301	302	303	304	305	306	307	308	309	310	311	312	313	314	315	316	317	318	319	320	321	322	323	324	325	326	327	328	329	330	331	332	333	334	335	336	337	338	339	340	341	342	343	344	345	346	347	348	349	350	351	352	353	354	355	356	357	358	359	360	361	362	363	364	365	366	367	368	369	370	371	372	373	374	375	376	377	378	379	380	381	382	383	384	385	386	387	388	389	390	391	392	393	394	395	396	397	398	399	400	401	402	403	404	405	406	407	408	409	410	411	412	413	414	415	416	417	418	419	420	421	422	423	424	425	426	427	428	429	430	431	432	433	434	435	436	437	438	439	440	441	442	443	444	445	446	447	448	449	450	451	452	453	454	455	456	457	458	459	460	461	462	463	464	465	466
---	---	---	---	---	---	---	---	---	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----

US *Acidothermus cellulolyticus*.

XH FH FT FT FT FT FT FT FT FT FT FT XX PN XX PD XX PF XX PR XX PA XX PI XX DR DR XX PT PT PT PT XX PS XX CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC

[illegible]

```
Query Match          7.9%; Score 180.4; DB 24; Length 3004;
Best Local Similarity 64.8%; Pred. No. 3e-25;
Matches 284; Conservative 0; Mismatches 151; Indels 3; Gaps 1
```

Qy	1852	CCAACCCCGTCCCGAGCCGAGCCGACCCCAAGCCCGAGTCTCTCCGCATCCCCGTC	1911
Db	2078	CCGAGCCGTGGCGAGTCGGACGCCGACCTACTCCGACCCGACACCGCCGACG	2137
Qy	1912	CCGAGCCCGACGCCAGCCCTTAGCCGTCGCCGTCCGCGAGCCGTCGCCGCGAGT	1971
Db	2138	CCAACGCTGACCCCTACTGCTACGCCACGCCACGGCAAGCCGACGCCGTACCGA	2195
Qy	1972	GTGTCGTCGTCGGGTGTGGGTGCCGGCGGAGGTATGTGGTGAATAGTGAATTGGG	2031
Db	2196	-CGCAGCCCTCCGAGGCCGCTGCACCGCGAGTTACACGGTCAACAGCGATTGGGCAAT	2254

QY 2032 GGCTTTACGGCAGCGTGGAGTACGAATACCGGAGCGCGGCGGACGCGGTGGACG 2091
 II III III III III III III III III III III III III III III III
 Db 2255 GGCTTACGGTACGCGTGGCGGTGACAAATTCGGATCCGTCGGACCAAGACATGGACG 2314
 QY 2092 GTGCGCTGGTCTGTTGGTGGGAATACAGCGGTACAGAACTACTGGACACTGGTTGACG 2151
 II III III III III III III III III III III III III III III III
 Db 2315 GTCACTGGACATTCGCGCGGAATACAGACGATTCACCAATTCGTGGAAATGCGAGCGGTCA 2374
 QY 2152 CAATCAGGTGCATCGGTGACGGCGGACGACCTGAGCTACAAACAGCTGATCCAAACCGGT 2211
 II III III III III III III III III III III III III III III III
 Db 2375 CAGAAGCGTCAGTCGGTACGGCTCGGAATATGAGTTATACAAACGATGATTCAGCGTGGT 2434
 QY 2212 CAGTCGACCACTTCGGATTCACGGAAGTTACTAGGAACAACACCGCACCTACACTC 2271
 II III III III III III III III III III III III III III III III
 Db 2435 CAGAACACCACTTCGGATTCACGCGGAGCTATACCGGAACCAACGCGGACCGACAGTC 2494
 QY 2272 ACCTGCACCGCTAGTTCA 2289
 II III III III III III
 Db 2495 GCCTGGCAGCAAGTTAA 2512

RESULT 4

AAX55661
 ID AAX55661 standard: DNA; 11706 BP.

XX
 AC AAX55661;

XX
 DT 30-JUL-1999 (first entry)

XX
 DE DNA sequence encoding truncated cellulases Cel B4/5 and Cel B5.

XX
 KW Cellulase; proteinase; truncated; Cel B5; Cel B4/5; Cel E1; Cel E1/2;
 KW Cel 1/2/3; Cel 6; Cel E3/B5; laundry detergent; stain-removing;
 KW cotton-containing fabric; stonewashing; ss.

OS Unidentified.

XX
 PN EP921188-A2.

XX
 PD 09-JUN-1999.

XX
 PF 15-SEP-1998; 98EP-0810919.

XX
 PR 19-SEP-1997; 97US-0932571.

XX
 PA (CLRN) CLARIANT FINANCE BVI LTD.

XX
 PI Anderson P, Bergquist PL, Daniels RM, Farrington GK;

PI Gibbs MD, Morgan H, Williams DP;

XX
 DR WPI: 1999-315403/27.

DR P-PSDB: AAY13492.

XX
 PT New truncated cellulase proteins, useful in detergents and for

PT producing 'stonewashed' denim

XX
 PS Disclosure: Page 20-23; 65pp; English.

XX
 CC The invention relates to a recombinant cellulase active protein free of
 CC proteinases of native thermophilic and alkaliphilic origin, comprising
 CC the truncated sequences Cel B5, Cel B4/5, Cel E1, Cel E1/2, Cel 1/2/3,
 CC Cel 6 or Cel E3/B5, or a stability region from one of the defined full-
 CC length sequences, or functional equivalents. Cel B5 extends from amino
 CC acid A1011 to P1424 or N1426, and Cel B4/5 extends from amino
 CC acid K635 to N1426 in the sequence shown in AAY13492; Cel E1 extends
 CC from amino acid Y39 to D481, Cel E1/2 extends from Y39 to G635, Cel
 CC E1/2/3 extends from Y39 to G812, Cel E6 extends from amino acid V1233 to
 CC K1751 and the stability region extends from amino acid E482 to G635 in
 CC the sequence shown in AAY13493; Cel E3/B5 is shown in AAY13494. The new
 CC enzymes are useful in laundry detergent compositions to prevent or
 CC remove staining, backstaining or graying, for use on cellulosic
 CC materials including cotton-containing fabrics. They are especially useful

CC for preventing redeposition of colorant during stonewashing, and for
 CC processing of textiles where cellulose breakdown is required. The new
 CC truncated enzymes show reduced redeposition of dye compared to using
 CC non-truncated cellulase compositions.

XX
 SQ Sequence 11706 BP; 3828 A; 1994 C; 2994 G; 2890 T; 0 other:

Query Match 5.8%; Score 133.6; DB 20; Length 11706;
 Best Local Similarity 52.4%; Pred. No. 2.5e-16;
 Matches 395; Conservative 0; Mismatches 344; Indels 15; Gaps 4;

QY 1220 CAATCAGTACCGGCACATCCCTCCGCCGTGCGGAGCTCGAGTCCATCTCTGTCGCGGT 1279
 III III III III III III III III III III III III III III III III
 Db 3883 CAACACACCTGCACCGACATCAGCCGCCACACCGAGCCCAACAGTCACAGCAACCGCG 3942
 QY 1280 CTCGTCGCGGTCTCCGTGCGCGTCTCCGTGCGGCTCTCCGTGCGGCTCTCTGTCGCGG 1339
 III III III III III III III III III III III III III III III III
 Db 3943 CTCCAACGCGCGCGGACAGTACGGTTACTGTGACTCCGACACCAACCAACACCGCA 4002
 QY 1340 GCCCGTCTCCGTCTGTCGCCGTGTTCGG---GTGGGTGAAGGTGCAGTACAGAACA 1396
 III III III III III III III III III III III III III III III III
 Db 4003 CGCGACACCGACAGGACACCTGGCACGGGAAGTGGTTGAAGGTACTATACAAGAACA 4062
 QY 1397 ATGATTGCGCGCGGTGATACACAGATCAACCGGCTCTCCAGTTGCTCAATACGGGT 1456
 III III III III III III III III III III III III III III III III
 Db 4063 ATGAGACAAGTCGAGCACAAGTTCTATAAGCGCGTGTCTTAAGATAGTCAATGGAGCA 4122
 QY 1457 CGTCGTCGGTGGATTTCTCCACGGTGACGTCGCGGTACTGTTCCACCGCGGATGGTGGT 1516
 III III III III III III III III III III III III III III III III
 Db 4123 GCAGCAGTGTGATCTTAGCAGGGTTAAGATAAGATACTGGTACACAGTGGATGGTG--- 4179
 QY 1517 CGTCGACACTGGTGTACAACCTGCACTGGCGGCGGTGGGTGTGGGAATATCCGCCCT 1576
 III III III III III III III III III III III III III III III III
 Db 4180 ACAAGCCACAGAGTCGGGTATGTGACTGGCCACAGATAGGGCAGCAATCTGACATTCA 4239
 QY 1577 CGTTGCGCTCGGTGAACCCCGGACCGGACGCGGACACTACCTGCAGTTGTGCTGTTCA 1636
 III III III III III III III III III III III III III III III III
 Db 4240 ATTTTGTGAAGCTGACGACGGGAGTCAGTGGAGCGGATATTACTTGGAGTAGGATTTA 4299
 QY 1637 CTGCTGGAAC-----GTTGGCGCGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 1690
 III III III III III III III III III III III III III III III III
 Db 4300 GCAGTGGAGTGGGCGAGTTGCACGCTGGTAAGACACACAGGGGATATACAGGTAAAGTTTA 4359
 QY 1691 ATAAGAGTACTGTCGAACCTTTGATGAGACCAATCTACTCTGATGGGACGAACACCG 1750
 III III III III III III III III III III III III III III III III
 Db 4360 ACAAGAATGACTGGAGCAATTTACAATCAGCAGACGACTGGTCATGCTTCAGAGCATGA 4419
 QY 1751 CCTTCCAGGATTGGA---CGAAGTGACGCTGATGTCATGCGCGGCTGGTGTGGGGGA 1807
 III III III III III III III III III III III III III III III III
 Db 4420 CGAATTATGGAGAGAAATGCGAAGTTAACGCTGTATGTAGATGCTGTCTGTATGGGGGC 4479
 QY 1808 CTGAACGTCGCGGACCGCCACAGCCCGCCACACCGCCCGCCAGCCCAACCCCGTCCCGCA 1867
 III III III III III III III III III III III III III III III III
 Db 4480 AGAGCGGGGCGGAGCGACACCTGCCACCGACAAGCAGCAGCAACCAACCGCAACTCCGA 4539
 QY 1868 GCCCGAGCGCGACCCCAAGCCCGAGCTCTTCCCATCCCGCTCCCGAGCGCCAGCCCGCA 1927
 III III III III III III III III III III III III III III III III
 Db 4540 CAGCAACAGCAACACCGCGCGCACAGCAACGCGCAACGCTCTACACCGCAGCAGCAGCA 4599
 QY 1928 GCCTACGCGCTCCCGTCCCGGAGCGCGCTGCGCC 1961
 III III III III III III III III III III III III III III III III
 Db 4600 CACCAACCCCAATPACCAACACCCCAACCGCTCC 4633

RESULT 5

AAD26525

ID AAD26525 standard: DNA; 11707 BP.

XX
 AC AAD26525;

XX
 DT 26-MAR-2002 (first entry)

XX
 DE Active cellulase protein, celB gene.

XX	Active cellulase protein; alkalophilic; textile processing; proteinnase;
KW	detergent additive; stonewashed appearance; cotton-containing denim;
KX	CelB5; thermophilic; commercial detergent; celB gene; ds.
XX	Unidentified.
OS	
FF	Location/Qualifiers
XX	6255..10535
FH	/tag= a
FT	/product= "CelB protein"
FT	/misc_feature
FT	8601..10532
FT	/tag= b
FT	/product= "DNA encoding B4/5 protein"
FT	9255..10526
FT	/tag= c
FT	/product= "DNA encoding B/5 protein"
XX	
XX	US6294366-B1.
PN	
XX	
PD	25-SEP-2001.
XX	
PD	19-AUG-1998; 98US-0136574.
XX	
PR	19-SEP-1997; 97US-0932571.
XX	
PA	(CLRN) CLARIANT FINANCE BVI LTD.
XX	
PI	Farrington GK, Anderson P, Bergquist P, Daniels R, Gibbs MD;
PI	Morgan H, Williams DP;
DK	WIPI: 2002-081780/11.
DR	P-PSDB: AAE16323.
XX	
PT	New cellulase active protein, useful in textile processing or
PT	commercial detergents, e.g. for improving the feel or appearance of
PT	cotton-containing fabrics, is stable under conditions of alkaline pH
PT	and elevated temperatures -
XX	
PS	Disclosure: Column 27-38; 6lpp: English.
XX	
CC	The present invention relates to a cellulase active protein, which is
CC	substantially free of proteinases of native thermophilic and
CC	alkaliphilic origin, where the cellulase active protein consists of the
CC	CelB5 amino acid sequence. The cellulase active protein is useful for
CC	treating cellulosic materials including cotton-containing fabrics, as
CC	detergent additives. The cellulase active protein is also useful for
CC	improving the feel and/or appearance of cotton-containing fabrics, for
CC	removing surface fibers from cotton-containing knits or for imparting
CC	stonewashed appearance to cotton-containing denims. The present proteins
CC	are stable under condition of alkaline pH and elevated temperatures,
CC	thus suitable for textile processing and in commercial detergents.
CC	The present sequence is celB gene.
XX	
SQ	Sequence 11707 BP; 3827 A; 1995 C; 2994 G; 2891 T; 0 other;
XX	
Query Match	5.8%; Score 133.6; DB 24; Length 11707;
Best Local Similarity	52.4%; Pred. No. 2.5e-16;
Matches 395; Conservative	0; Mismatches 344; Indels 15; Gaps 4
OY	1220 CAATCAGTACCGGCACATCGCTCCGCCGTCGGCACGTCACTCTCGCCGT 1279
Db	3883 CAACCACACTTGCCACCGACATCAGCCCGACACCGACCCCAACAGTGTGAGCAACGCCGA 3942
OY	1280 CTCCTCGCCGCTCTCCGTCGGCGTCTCCGTCGGCGTCTCGTCCGCTTCGTCGCCGA 1339
Db	3943 CTCCAAACCGCCGACCGACATGACGGTTACTGTGACTCCGACACCCACACCAACACCCA 4002
OY	1340 GCCCTCTCTCGTCTCGTCGGCGGTGTCCG---GTGGGGTGAAGGTGCAGTACAAGAACA 1396
Ref.	4003 C*CCGACACCGACGAGGACACCTGCCACGGGAGCTGGTTTGAAGGTACTATACAAGAACA 4062
 CCUUGGGTCATTAACCATGATCAAACCGGGTCTCCAGTTGGTGAATAGGGGT 1456

QY 841 TGGCCATACAAAGACCCGTCGACGGCGTGGACAAACAATGCTCTTCTCCGTGTCAAGAAC 900
DB 742 TGGCCCTACAACGGCTGGTC---CGGTGTGACTGCAAGAAAGCTCCTTTCGATAGAGACG 798
QY 901 ATTGACTTTGGCAGCTATCACCTGTACCCCAATTACTGGGG-----CCAGAAC 948
DB 799 GTGGACTTCGGCAGCTTCCACCTCTATCCGTCCCACTGGGGTGTCACTCCAGAGAATCTAT 858
QY 949 GCGGACTTGGGAACGCAATCGATCAAGGATCATATTGCCGAATGCCGACGCGATCGGCAAG 1008
DB 859 GCCCAGTGGGAGCGAAGTGGATAGAGACCACATAAAGATCCAAAGAGATCGGAAA 918
QY 1009 CCGACCATCTCTGAAATTCGG-----CTGGCAGACACCGGACCGGATTCGGTC 1059
DB 919 CCGCTTTGTTGGAAGAATATGGAATTCCAAAGAGTCCGCCAGTTAACAGACGGCCATC 978
QY 1060 TATCAGACGTGGACCCAGACTGTGCGCTAGCAACGGGTGAACGCGCTGGAACCTTCTGGATG 1119
DB 979 TACAGACTCTGACACGATCTGGTCTAGGATCTCGGTGGAGATCGGAGGATGTTCTGGATG 1038
QY 1120 CTGCTGGGAATGTCAACGG 1139
DB 1039 CTCGCGGAATCGGGAAG 1058

RESULT 10
AAT94213
ID AAT94213 standard; DNA: 2043 BP.
AC AAT94213;
XX
DT 21-MAY-1998 (first entry)
DE Thermotoga maritima endoglucanase DNA.
XX
KW Endoglucanase; cellulase; carboxymethylcellulose; cellulose;
KW biomass; beta-1,4-glycosidic bond; hydrolysis; saccharification;
KW thermostable enzyme; thermophilic; glycosidase; ss.
XX
OS Thermotoga maritima MSB8 (clone 6GP2).
XX
PN WO9744361-A1.
XX
PD 27-NOV-1997.
XX
PF 22-MAY-1997; 97WO-US08793.
XX
PR 22-MAY-1996; 96US-0651572.
XX
PA (RECO-) RECOMBINANT BIOCATALYSIS INC.
XX
PI Lam DE, Mathur EJ;
XX
XX WPI: 1998-018435/02.
DR P-PSDB; AAW35005.
XX
PT Endoglucanase(s), preferably form archaeal bacterium, AEP11 la -
PT useful to degrade carboxymethylcellulose and hydrolyse of
PT beta-1,4-glycosidic bonds in cellulose
PS Claim 3; Fig 10; 164pp; English.
XX
XX This DNA sequence from Thermotoga maritima MSB8 (clone 6GP2)
CC encodes an endoglucanase (see AAW35005) that is able to degrade
CC carboxymethylcellulose and to hydrolyse the beta-1,4-glycosidic
CC bonds in cellulose, and which shows homology to the thermostable
CC endoglucanase (see AAW34985) of archaeobacterium hydrothermal vent
CC isolate AEP11a. The DNA can be used in the recombinant production
CC of the endoglucanase and as a probe to identify similar sequences.
CC 24 Endoglucanase polynucleotides (see AAT94193-216) are claimed.
CC These can be incorporated into plasmid or virus-derived vectors for
CC use in a claimed method of producing enzymes in transformed host

CC cells. The claimed endoglucanases (see AAW34985-W35008) can be used
CC to degrade cellulose for the conversion of plant biomass into fuels
CC and chemicals, for use in detergents, textiles, animal feed, waste
CC treatment, and in the fruit juice and brewing industries for the
CC clarification and extraction of juices.
XX
SQ Sequence 2043 BP; 609 A; 417 C; 592 G; 425 T; 0 other;
Query Match 5.5%; Score 126.4; DB 19; Length 2043;
Best Local Similarity 49.3%; Pred. No 51e-15;
Matches 513; Conservative 0; Mismatches 476; Indels 51; Gaps 5;
QY 121 TTCGTCCACCOCATCCGGCGGTCTAGTTCTTCAACCGCCCTTCCCTATCTGTTACGGGGA 180
DB 100 TTCGTGAACTGCAAAACGCAAAATTCGCTCTCAACGGAAGAATTCAGATTCATTGGA 159
QY 181 ACGAACAACATATTACCTCAGCTATCATGTCGACGCGCCGACGTCGATGACGTGTTGGCCAA 240
DB 160 AGCAACAACATCTACATGTCACATGCAAGACGACGGAATGATAGACAGTGTTCGGAGAGT 219
QY 241 GCTCAACGCGATGAATCTTTCTGTATCCGCGACCTGGGGTTTTCATCGACATCGGCTCTTT 300
DB 220 GCCAGAGACATGGGTATAAGGTCTCTCAGAAATCTGGGGTTTCTCGACGGGAGAGTTAC 279
QY 301 GACGGTCTCGTCCCAACATTCGATGGCAACAAGACGGCTTCTACTTTTCAGTACTGGAC 360
DB 280 TGCAGAGACAAGAACACCTACATGTCATCTGAGCCCGGTCTTTTCGGGGTCCCAAGAGGA 339
QY 361 CCGTCGACCGCGCTCCGGCTACACGCGCGCCGACCGGCTTGAAGGCCCTTGACTAC 420
DB 340 ATATCGAACGCC-----CAGAGCGGTTTCGAAAGACTCGGACTAC 378
QY 421 CGCATCCGACGCGCGCGCGCGCTTCCGGTGTATGTCGCTCTCACCACGACTGG 480
DB 379 ACAGTTGCGAAACGGAAGAACTCGGTATATAAACTTCTCATTTCTTCTGAACAACCTGG 438
QY 481 AAGAATTTGGGGGAATGGATCAATACGCAAGTGGTACGGCTTCTTACCACGACACAC 540
DB 439 GAGGACTTCGGTGAATGAACCACTAGTACGTGAGGTGGTTGGAGGAACCCATCAGCAGCAT 498
QY 541 TTCTACACGACCCCGGACCGCGTACAGAATTTGGGTCAATCATCTACTCAAC 600
DB 499 TTCTACAGAGTGAAGAAGATCAAGAAGAGATCAAAAAGTACGCTCTCTTCTCGTAATAC 558
QY 601 CGGCTCAACAGCATTTACCGCGGTGAGTACAAAGAACGATCCAAAGCTTTTGTCTGGAA 660
DB 559 CATGTCAATACCTACACGGGAGTTCCTTACAGGAAGAGCCACCATCATGCGCTGGGAG 618
QY 661 CTTGCCAATGAGCCGCGTGGTAGGAAGCGGCACATTTACCAACCTCGGCGACGTCGACT 720
DB 619 CTTGCAACGCAACCGCGCTGTGAGACGGCAAAATCGGGGAACACGCTCTGTAGTGGTG 678
QY 721 CAGGCGACCATTTGTCACCTGGGTGATCAAAATGTCGGCGTACGTCAAAAACATAGACCT 780
DB 679 AAGGAGA-----TGAGTCTCTACATAAAGAGTCTGGATCCCAACACCTCGTGGTGTG 732
QY 781 AACCATATGCTCTCGGTCCGCGACGAGGTTTCTACATTTGGGTCAACGCGAGGAAGCGG 840
DB 733 GGGGACGAAGGATTTCTTACGCAACTACGAAGGATTCAAACCTTACGGTGGAGAACCGGAG 792
QY 841 TGCCCATACAACGACCCGTCGACGCGGTGCGACAACATGCTCTTCTCCGTGTCAAGAAC 900
DB 793 TGGGCTTACAACGGCTGGTCT---CGGTGTGACTGGAAGAAGCTCTTTTCGATAGAGACG 849
QY 901 ATTGACTTTGGCAGCTATCACCTGTACCCCAATTACTGGGG-----CCAGAAC 948
DB 850 GTGGACTTCGGCAGCTTCCACCTCTATCCGTCCCACTGGGGTGTGACTGCAGAGAATCTAT 909
QY 949 GCGGACTGGGGAACGCAATGGATCAAGGATCATATTGCCGAATGCCGACGCGATCGGCAAG 1008
DB 910 GCCCAGTGGGAGCGGAAGTGGATAGAGACCACATAAAGATCGCAAAAGAGATCGGAAA 969
QY 1009 CCGACCATTTCTCGAAGAATTCGG-----CTGGCAGACACCGGACCGGATTCGGTC 1059

Db	970	CCCGTTGTTCTGGAGAATATGGAATTCAAAGAGTGCGCCAGTTACAGAACGCCGATC	1029
Qy	1060	TATCAGACGTGACCCAGACTGTGGCTACGACCGCTGAAGCAGGCTGGAACCTCTCGGATG	1119
Db	1030	TACAGACTCTGGACGATCTGGTCTACGATCTCGGTGGAGATGAGCGGATCTCTCGGATG	1089
Qy	1120	CTCGCTGGGAATGTCAACGG	1139
Db	1090	CTCGCGGAATCGGGAAGG	1109

RESULT 11

AAV36917
ID AAV36917 standard; DNA: 2043 BP.

AA
AC AAV36917;

DT 21-DEC-1998 (first entry)

DE Thermotoga maritima 6GP2 beta-mannanase gene coding region.

KW Glycosidase; 6GP2; thermostable enzyme; oligosaccharide; glucose;
 KW sugar; baking; textile; detergent; beta-mannanase; ss.

XX OS Thermotoqa maritima strain 6GP2.

PN WO9824799-A1.

11-JUN-1998.

08-DEC-1997;

PR 10-OCT-1997; 97US-0949026.

XX
XX

XX
XX

[illegible]

DR P-PSDB; AAW49868.

PT Glycosidase enzym

PT
PT
PT
PT

PS Claim 1; Flg 11a-d; 92pp; English.

this isolated polynucleotide comprises a coding region for
 beta-mannanase 6G82 (see AAW49868) from a *Thermotoga maritima*
 clone (6GP2) that grows optimally at 85 degC in high salt medium.
 The sequence shows 37% nucleic acid identity to a mannanase of
Aspergillus aculeatus. The invention provides 18 polynucleotides
 (see AAV36907-24) coding for the most stable glycosidases (see AAW49858-75)
 having glucosidase, alpha-galactosidase, beta-galactosidase,
 beta-mannosidase, beta-mannanase, endoglucanase or pullulanase
 activity. Vectors and host cells are also claimed. A method is
 provided for producing the enzymes by recombinant techniques. A
 claimed method for generating glucose from soluble cell
 oligosaccharides comprises contacting a sample (selected from dairy
 products, fruit juice, detergent, textile, guar gum, animal feed,
 plant biomass or waste product) containing oligosaccharides
 (selected from maltose, cellobiose, lactose, sucrose, raffinose,
 stachyose, verbascose, cellulose, starch, amylose, glycogen,
 disaccharides, polysaccharides and pullulan) with one of the
 claimed glycosidases such that glucose is produced.

SQ Sequence 2043 BP; 610 A; 419 C; 589 G; 425 T; 0 other;

Query Match

ISSN 0013-788X; FLEU. NO. 1E-14;

Matches	512:	Conservative	0:	Mismatches	477:	Indels	51:	Gaps
Qy	121	TTCTGTCACCGCATCCGGCGGTCAGTTCTTCTGTAACGGCCCTTCCCTATATCGTTACGGGGGA	180					
Db	100	TTCTGTAAGTGGAAACCGAAATTCCTCTGAACGGAAAGAATTCAGATTTCATTGGA	159					
Qy	181	ACGAACAACATATTACCTCAGCTATCAGTCGACGCGGACGTCGATGACGTGTGGCCAAAG	240					
Db	160	AGCAACAACATATTACATCATGCTACAAAGACCAACGGAATCATAGACAGTGTCTCGAGACT	219					
Qy	241	GCTCAAGCGATGAATCTTTCTGTCTATCCGACCTTGGGGTTTTCATCGACATCGGCTCTCTT	300					
Db	220	GCCAGACATCGGGTATTAAGGTCCTCAAGATCTGGGGTTTTCCTCGACGGGAGAGTTTAC	279					
Qy	301	GACGGCTCCGTGCCACAAATCATGCGCAACAGAACGGCTTCTACTTTTCAGTACTTGGGAC	360					
Db	280	TGCAGACACAAGAACACCTACATCATCTGAGCCCGGTGTTTCGGGGTCCACAGAAGGA	339					
Qy	361	CGGTGACCGGGCTTCGGGCTTACAACGACGGGCGGACGGCTTGTCAAGGCTTTGACTTAC	420					
Db	340	ATATCGAACGGC-----CAGAGCGGTTTCGAAGAAGACTCGACTTAC	378					
Qy	421	GCGATCGGAGCGGGCGCGCACGGCTTCGGGTGATTGTTCGTCTTCCACCAACGACTGG	480					
Db	379	ACAGTTGCGGAAGCGAAGAACTCGGTATAAACTTGTTCATTGTTCTTGTGAACAACACTG	438					
Qy	481	AAAGAAATTGGGGGAATTCGATCAATACCACAAAGTGGTACGGCTTCTCTTACCAACGACAAC	540					
Db	439	GACGACTTCGGTGGAAATCAACCACTACGTGAGTGGTTTGGAGGAAACCATTCACGACGAT	498					
Qy	541	TTCTACACCGACCCCGGACGCGGTACAGAAATTTGGGTCAATCATCTACTGAAC	600					
Db	499	TTCTACAGAGATGAGAAGATCAAAAGAGATACAAAAGTACGTCTCTCTTCTCTGTAAC	558					
Qy	601	CGGTTCACACGATTAACGGCGTGACGTACAAGAACGATCCAACGATCTTTCCTTGGGAA	660					
Db	559	CATGTCAATACCTACACGGGAGTTCTTTTACAGGGAAGACGCCACCATCATGGCTGGGAG	618					
Qy	661	CTTGCATGAGCGCGCTGCGTAGGAAGCGGCACATTTACCAACCTCGCGGACGTCGACT	720					
Db	619	CTTGCAACGAACCGCCCTGTGACAGCGCAAAATCGGGGAACAGCTCGTTGAGTGGGTG	678					
Qy	721	CAGCGCACCATTTGTCACTGGGTGATCAATGTGCGGTCAGTCAAAAGCATAGACCT	780					
Db	679	AAGGAGA-----TGAGCTCTACATAAAGAGTCTGGATCCCAACCACTCTGTCGTG	732					
Qy	781	AACCATATGTTCTCGGTGCGGACGAAGGTTTACATTTGGGTCAACCCAGGGAAGCGGC	840					
Db	733	GGGACGAAGGATTTCTCAGCAACTACGAAGGATTCAAACCTTACGGTGGAGAAGCCGAG	792					
Qy	841	TGGCCATACAAGACCCGTCGACGGCTCCACAAACATGCTCTTCTCCGTGTCGAAGAC	900					
Db	793	TGGGCTACACGGTGGTCT---CGGTGTTGACTTGGGAAGAGCTCTTTTCGATAGACAG	849					
Qy	901	ATTGACTTTGCGCATATCACCTGTACCCGAATTTACTGGGG-----CCAGAAC	948					
Db	850	GTGACTTCGGCACGTTCCACCTCTATCCGTCCCACTGGGGTGTGCTCAGACAACTAT	909					
Qy	949	GCGGACTGGGAACCGCAATGGATCAAGGATCATATTTCGGAATGCGGACGCGATCGGCAAG	1008					
Db	910	GCCAGTGGGAGCGAAGTAGATAGAAGACCATATAAGATTCGCAAAAGAGATCGGAAA	969					
Qy	1009	CCGACCATTCTCGAAGAAATTCGG-----CTGGGACACACCGACCCCGATTCCGTC	1059					
Db	970	CCCGTTGTTCTGGAAGAATATGGAAATTTCCAAAGAGTCCGCCAGTTTACAGAACCGCCATC	1029					
Qy	1060	TATCAGACGTGGACCCAGACTGTGCGTACGAACCGTGAAGCAGGCTGGAATCTTCTGGAT	1119					
Db	1030	TACAGACTCTGGAACGATCTGGTCTACGATCTCGTGGAGATGGAGCGATGTTCTGGATG	1089					
Qy	1120	CTCGCTGGGAATGTCAACGG	1139					
Db	1090	CTCGCGGAATTCGGAGAGG	1109					

RESULT 12
 AAT94200
 ID AAT94200 standard; DNA; 2001 BP.
 XX
 AC AAT94200;
 XX
 DT 21-MAY-1998 (first entry)
 XX
 DE Thermotoga neapolitana endoglucanase DNA.
 XX
 KW Endoglucanase; cellulase; carboxymethylcellulose; cellulose;
 KW Biomass; beta-1,4-glycosidic bond; hydrolysis; saccharification;
 KW thermostable enzyme; thermophilic; glycosidase; ss.
 XX
 OS Thermotoga neapolitana (clone 56Gp1).
 XX
 PN W09744361-A1.
 XX
 PD 27-NOV-1997.
 XX
 PF 22-MAY-1997; 97WO-US08793.
 XX
 PR 22-MAY-1996; 96US-0651572.
 XX
 PA (RECO-) RECOMBINANT BIOCATALYSIS INC.
 XX
 PI Lam DE, Mathur EJ;
 XX
 PF WPI: 1998-018435/02.
 DR
 DR P-PSDB; AAW34992.
 XX
 PT Endoglucanase(s), preferably form archaeal bacterium, AEP11 la -
 PT useful to degrade carboxymethylcellulose and hydrolyse of
 PT beta-1,4-glycosidic bonds in cellulose
 XX
 PS Claim 3; Fig 1H; 164pp; English.
 XX
 CC This DNA sequence from Thermotoga neapolitana (clone 56Gp1) encodes
 CC an endoglucanase (see AAW34992) that is able to degrade
 CC carboxymethylcellulose and to hydrolyse the beta-1,4-glycosidic
 CC bonds in cellulose, and which shows homology to the thermostable
 CC endoglucanase (see AAW34985) of archaeobacterium hydrothermal vent
 CC isolate AEP11la. The DNA can be used in the recombinant production
 CC of the endoglucanase and as a probe to identify similar sequences.
 CC 24 Endoglucanase polynucleotides (see AAT94193-216) are claimed.
 CC These can be incorporated into plasmid or virus-derived vectors for
 CC use in a claimed method of producing enzymes in transformed host
 CC cells. The claimed endoglucanases (see AAW34985-W35008) can be used
 CC to degrade cellulose for the conversion of plant biomass into fuels
 CC and chemicals, for use in detergents, textiles, animal feed, waste
 CC treatment, and in the fruit juice and brewing industries for the
 CC clarification and extraction of juices.
 XX
 SQ Sequence 2001 BP; 598 A; 394 C; 580 G; 429 T; 0 other;
 Query Match 5.4%; Score 124.4; DB 19; Length 2001;
 Best Local Similarity 53.0%; Pred. No. 1.2e-14;
 Matches 424; Conservative 0; Mismatches 316; Indels 60; Gaps 5;
 QY 376 CCGCGGTACAAACACCGCGCGCGCTTCCAGGCGCTTCACTACGGATCGGAGCGCG 435
 Db 298 CCAGAGGTCAGAACCGCTCAGACGGTTTGAAGACTCGCACTACAGGTAGCGAAGCA 357
 QY 436 GCGCGCACCGCGCTTCGGGTGATTGTCTCCACCAACCACTGGAAGAAATTGGGGGA 495
 Db 358 AAAGAACTGGGCATCAAGCTCAATCGTTCTTGTGNACAACTGGGACGATCGGTGGA 417
 QY 496 ATGATCAATACACAAAGTGTACCGGCTTCTTACACGACAACTTCTACACCGACCCC 555
 Db 418 ATGAATCAATACGTGATGATGTTGGGGGCATCCATCAGCATG:..TCTACAGGAACGAG 477

QY 556 CGACCCAGCAGCGGTACAAGAATTGGGTCAATCATCTACTGAACCGGTCAACAGCAATT 615
 Db 478 AAGATCAAGAAGAATAACAAAAGTACGTGCTTTCCTCATATAACAGGGTGAACACTTAC 537
 QY 616 ACCGGGTGACGTACAAGAAGCATCCACAGACTCTTTGGTGGGAATTTGCCAATAGCGG 675
 Db 538 ACGGTGTTCTTACAGGGAAGAGCCACCATCATGTGGTACGGAATGCGGAACGAGGCC 597
 QY 676 CGCTCGGTAGGAAGCGGCACATTACCAACCTCGGGCAGCTGCACCTCAGGCGACCAATTGC 735
 Db 598 AGTGTGAACGGAC-----AAGTCTGGTAAACACACTCGTT 633
 QY 736 AACTGGGTGATCAAAATGTGCGCGTACGTCAAAAGCATAGACCCCTAACCATATGCTCTCG 795
 Db 634 GAATGGGTAGAGAGATGAGTGCTTACATAAAGAGTCTGGATCCAAACCACTGGTGGC 693
 QY 796 GTCGGGACGAAGGGTTCTACA-----TTGGGTCAACGAGGGAACCGGCT 841
 Db 694 GTGGGAGACGAGGATTCTTCAACAACACTACGAAGGCTTTCAGACCTTACGGTGGAGAGCT 753
 QY 842 G-GCCATACAACGACCGCTCCGACGCGTCCGACACAACTTACTTCTCGGTCTCAAGAAC 900
 Db 754 GAGTGGGCTTACAACGGATGTCGCGTGTGACTGGAAGAGACTTCTGGAGATAGAGACG 813
 QY 901 ATTGACTTTTGGCAGGTATCACTGTACCCGAATTACTGGGCCAGA-----AC 948
 Db 814 GTGATTTTGGTACGTTCCTCTCTACCCCTCCCACTGGGGTGTGAGCGCTCAAACTAC 873
 QY 949 GCGACTGGGGAACGAATGATCAAGGATCATATTGGCAATGCGGCGCAGGATCGGCAAG 1008
 Db 874 GCACAGTGGGGGAAAGTGGATAGAGATCACAATAAAGATCGCAAAAGAGGTTGCAAAA 933
 QY 1009 CCGACCAATTCGGAAGATTTCGGCTGGCAGACACCGG-----ACGGCATTCGGTC 1059
 Db 934 CCCGTGTTCTGGAAGAGTACGGTATTCCCAAAAGTCCCGCGCTCAACAGGGTTCCCAAT 993
 QY 1060 TATCAGACGTGGACCCGACACTGTGCGTACGAACGAGGCTGGAACCTTCTGGATG 1119
 Db 994 TACAAATTTGGAACGATCTGCTCTACAACTCGTGGAAACGGTGGCCATGTTCTGGATG 1053
 QY 1120 CTCCTCGGGAATGTCAACGG 1139
 Db 1054 CTCGAGGAATCGGTGAAGG 1073
 RESULT 13
 AAW13001
 ID AAW13001 standard; DNA; 2977 BP.
 XX
 AC AAW13001;
 XX
 DT 14-OCT-1991 (first entry)
 XX
 DE Endol gene encoding endoglucanase.
 XX
 KW Cellulase activity; detergent; ds.
 XX
 OS Bacillus spp. NCIMB 40250.
 XX
 FH Key Location/Qualifiers
 FT CDS 677..2779
 FT sig_peptide /*tag= a
 FT mat_peptide /*tag= b
 FT 677..769
 FT 770..2776
 PN W09110732-A.
 XX
 PD 25-JUL-1991.
 XX
 PF 18-JAN-1991; 91WO-DK00013.
 XX
 PR 19-JAN-1990; 90DK-0000164.

Db	1502	AGCCCCACACAGATTTCGGCTTCCTCGGCTCTTAAGGGCAACTCCAACCTCTGTTCGACCCCT	1501			
Qy	2271	CACCTGCACGGCTAGTTGA	2289			
Db	1562	TACCTGGCGCCGACGCTGA	1580			
RESULT 15						
AA	AA	AA	AA			
ID	AA	AA	AA			
AC	AA	AA	AA			
AC	AA	AA	AA			
XX	30-JUL-1999	(first entry)				
XX	DNA sequence encoding truncated cellulase Cel E3/B5.					
DE	Cellulase: proteinase; truncated; Cel B5; Cel B4/5; Cel E1; Cel E1/2;					
KW	Cel 1/2/3; Cel 6; Cel E3/B5; laundry detergent; stain-removing;					
KW	cotton-containing fabric; stonewashing; ss.					
XX	Unidentified.					
OS						
XX	EP921188-A2.					
XX	09-JUN-1999.					
XX	15-SEP-1998:	98EP-0810919.				
XX	19-SEP-1997:	97US-0932571.				
XX	(CLRN) CLARIANT FINANCE BVI LTD.					
PA						
PI	Anderson P, Bergquist PL, Daniels RM, Farrington GK;					
PI	Gibbs MD, Morgan H, Williams DP;					
XX	WP1; 1999-315403/27.					
DR	P-PSDB; AAV13494.					
XX						
PT	New truncated cellulase proteins, useful in detergents and for					
PT	producing 'stonewashed' denim					
XX	Disclosure; Page 41-42; 65pp; English.					
XX						
CC	The invention relates to a recombinant cellulase active protein free o					
CC	proteinases of native thermophilic and alkaliphilic origin, compris					
CC	the truncated sequences Cel B5, Cel B4/5, Cel E1, Cel E1/2, Cel 1/2/3,					
CC	Cel 6 or Cel E3/B5, or a stability region from one of the defined fu					
CC	length sequences, or functional equivalents. Cel B5 extends from amino					
CC	acid A1011 to P1424 or K1425 or N1426, and Cel B4/5 extends from amino					
CC	acid K635 to N1426 in the sequence shown in AAV13492; Cel E1 extends					
CC	from amino acid Y39 to D481, Cel E1/2 extends from Y39 to G635, Cel					
CC	E1/2/3 extends from Y39 to G812, Cel E6 extends from amino acid V1233					
CC	K1751 and the stability region extends from amino acid E482 to G635 in					
CC	the sequence shown in AAV13493; Cel E3/B5 is shown in AAV13494. The new					
CC	enzymes are useful in laundry detergent compositions to prevent or					
CC	remove staining, backstaining or graying, for use on cellulosic					
CC	materials including cotton-containing fabrics. They are especially use					
CC	for preventing redeposition of colorant during stonewashing, and for					
CC	processing of textiles where cellulose breakdown is required. The new					
CC	truncated enzymes show reduced redeposition of dye compared to using					
CC	non-truncated cellulase compositions.					
XX						
SQ	Sequence 2029 BP; 650 A; 329 C; 520 G; 530 T; 0 other;					
Query Match						
Best Local Similarity 3.9%; Score 90.2; DB 20; Length 2029;						
Matches 306; Conservative 0; Mismatches 268; Indels 15; Gaps						
Qy	1370	GTGGGTGAAGTGCAGTACAGAACATGATTCGGCGCCGGGTGATTAACCATCAAC	1429			
Db	8	GTGGTGTGAAGTACTGTACAGAACATGAGAACGTGCAGCACAGGTTCTATAAGGC	67			

GenCore version 5.1.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 13, 2002, 08:52:51 ; Search time 2135 Seconds
(without alignments)
17363.686 Million cell updates/sec

Title: US-09-917-378-2

Perfect score: 2289

Sequence: 1 atgddgtactgctgcgcgc.....tcacgtgcacgctagtgtga 2289

Scoring table: IDENTITY_NUC

Gapop 10_0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:*

1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estm:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_hic:*
9: gb_est1:*
10: gb_est2:*
11: gb_hic:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: gb_gss:*
18: em_gss_hum:*
19: em_gss_inv:*
20: em_gss_pln:*
21: em_gss_vrt:*
22: em_gss_fun:*
23: em_gss_mam:*
24: em_gss_mus:*
25: em_gss_other:*
26: em_gss_pro:*
27: em_gss_fro:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	81.2	3.5	974	17	BH157536	ENTRL88TR
2	80.8	3.5	974	17	BH157536	ENTRL88TR
3	79.4	3.5	335	17	A2415847	IM0190C20
4	79.2	3.5	463	13	BI781048	EBma03_SQ
5	79.2	3.5	678	14	BQ760233	EBp107_SQ
6	78.6	3.4	591	12	BG240228	OV1_19_DI

7	78.6	3.4	655	13	BJ248065	BJ248065
8	76.4	3.3	1315	10	AV752548	AV752548
9	76.2	3.3	788	17	CNS04AAV	AA720369
10	76	3.3	236	9	AA720369	33562 Lam
11	75.6	3.3	586	17	A2871772	A2871772
12	75.2	3.3	939	17	AG043613	AG043613 Pan trogl
13	74.4	3.3	829	17	CNS035GN	AL228704 Tetraodon
14	73.6	3.2	818	17	CNS044N4	AL274297 Tetraodon
15	72.4	3.2	616	17	A2432179	A2432179 IM0217A05
16	72	3.1	988	17	CNS022KP	AL178306 Tetraodon
17	71.2	3.1	135	9	AL820149	AL820149
18	71	3.1	955	17	CNS02RNO	AL210813 Tetraodon
19	70.8	3.1	865	12	BF315505	BF315505 601899637
20	70.6	3.1	162	12	BE918956	BE918956 FM1_2_C07
21	70.6	3.1	925	17	CNS00CAZ	AL058951 Drosoph11
22	70.2	3.1	464	12	BE799902	BE799902 60158045
23	70.2	3.1	622	13	BI994975	BI994975 1031018D0
24	69.6	3.0	562	17	A2855343	A2855343 2M0159A22
25	69.6	3.0	610	17	A2369884	A2369884 IM0120111
26	69.4	3.0	559	17	BH306809	BH306809 CH230-100
27	69.4	3.0	791	17	CNS028QT	AL186302 Tetraodon
28	69.2	3.0	724	17	A2197886	A2197886 SP_1035_B
29	69.2	3.0	2003	12	BG3303570	BG3303570 602430456
30	68.8	3.0	464	17	A2928926	A2928926 479.d1f19
31	68.2	3.0	438	17	A2858646	A2858646 2M0163J20
32	67.8	3.0	843	17	CNS00CS1	AL059666 Drosoph11
33	67.8	3.0	875	17	AG043475	AG043475 Pan trogl
34	67.6	3.0	826	17	BH316049	BH316049 CH230-40E
35	67.4	2.9	927	17	AG127811	AG127811 Pan trogl
36	67.4	2.9	955	17	CNS02YA4	AL219397 Tetraodon
37	67.2	2.9	527	17	A2297141	A2297141 RPCL-23-1
38	67.2	2.9	565	17	A2386138	A2386138 IM0144P14
39	67.2	2.9	959	17	CNS022KT	AL221078 Tetraodon
40	67	2.9	828	17	CNS04XH7	AL311668 Tetraodon
41	66.8	2.9	1160	17	AG043473	AG043473 Pan trogl
42	66.6	2.9	402	14	BQ568798	BQ568798 g1114h07
43	66.6	2.9	600	12	BG802543	BG802543 0182-32 M
44	66.6	2.9	804	13	BI946620	BI946620 HVSMF1000
45	66.6	2.9	976	17	A2668903	A2668903 ENTIC14TR

ALIGNMENTS

RESULT 1
BH157536
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

BH157536
ENTRL88TR Entamoeba histolytica Sheared DNA
genomic, DNA sequence.
BH157536.1 GI:15730974
GSS
Entamoeba histolytica.
Entamoeba histolytica
Eukaryota; Entamoebidae; Entamoeba.
1 (bases 1 to 974)
Loftus B., Wang Z., Van Aken S. and Fraser C.
Determination of clone end sequences from Entamoeba histolytica
HM1:IMSS sheared DNA library (2001)
Unpublished (2001)
Contact: Brendan J Loftus
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0208
Fax: 301 838 3543
Email: b.loftus@tigr.org
Clones are derived from the Entamoeba histolytica HM1:IMSS sheared
DNA library
Seq primer: M13-Reverse
Class: shotgun
High quality sequence start: 68
High quality sequence stop: 206.

U l:
oping
barley
ptome
stigmating

1872

108
1932
168
1992
228

phyta;
Pooldeae

COMMENT

Contact: Waugh R, Marshall DF
Genome Dynamics/Computational Biology
Scottish Crop Research Institute
Invergowrie, Dundee, DD2 5DA, Scotland, UK
Tel: 00 44 1382 562731
Fax: 00 44 1382 562426
Email: est@scri.sari.ac.uk.

FEATURES

source

1. .678
Location/Qualifiers
/organism="Hordeum vulgare"
/cultivar="Optic"
/db_xref="taxon:4513"
/clone="Epi107_SQ003_J13"
/clone_lib="pistil, 12 DPA, no treatment, cv Optic, Epi107"
/tissue_type="pistil"
/dev_stage="12 DPA"
/lab_host="DH10B"
/note="Vector: pSPORT1; Site_1: Sal I; Site_2: Not I;
Non-normalised library, directionally cloned into pSPORT1.
Derived from pistils dissected from developing grains (12
days post anthesis) in glasshouse grown barley plants.
Developed as part of the barley transcriptome resources of
BSRC/SEERAD funded cereal IGF (Investigating Gene
Function) project."

BASE COUNT

131 a 240 c 186 g 121 t

ORIGIN

Query Match 3.5%; Score 79.2; DB 14; Length 678;
Best Local Similarity 62.8%; Pred. No. 6.8e-06;
Matches 123; Conservative 0; Mismatches 73; Indels 0; Gaps 0;
QY 1813 CCGTCGGGACACCGCCAGCCACACCGCCAGCCCGCCAGCCCGGTCGGGCGCCG 1872
DB 286 CCGTGGCGGAGAGCGGAGCGCCAGCCAGCCCGCCAGCCAGCCAGCCGTCGCCA 345
QY 1873 AGCCGAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCC 1932
DB 346 AGCCGAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCC 405
QY 1933 ACCCGGTCGCCGCGGAGCCGTCGCCGTCGCCGTCGCCGTCGCCGTCGCCGTC 1992
DB 406 ACCCGGACTACACCGCGTCGCCGTCGCCGTCGCCGTCGCCGTCGCCGTCGCCG 465
QY 1993 TGGCGGCGGACGATG 2008
DB 466 ACCACGCGCCACCACTG 481

RESULT 6

BG240228

LOCUS

BG240228 591 bp mRNA linear EST 15-FEB-2001
DEFINITION OVI_19_D11.bl_A002 Ovary 1 (OVI), Sorghum bicolor cDNA, mRNA

ACCESSION

BG240228

VERSION

BG240228.1

KEYWORDS

EST.

SOURCE

Sorghum.

ORGANISM

Sorghum bicolor

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC

clade; Panicoideae; Andropogoneae; Sorghum.

1 (bases 1 to 591)

Cordonnier-Pratt,M.-M., Gingle,A., Marsala,C., Sudman,M. and Pratt

,L.H.

An EST database from Sorghum: ovaries of varying immature stages

Unpublished (2000)

Contact: Cordonnier-Pratt MM

Laboratory for Genomics and Bioinformatics

The University of Georgia, Department of Plant Biology

Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA

Tel: 706 542 1860

Fax: 706 583 0210

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

FEATURES

source

1. .591
Location/Qualifiers
/organism="Sorghum bicolor"
/db_xref="taxon:4538"
/clone_lib="Ovary 1 (OVI)"
/note="Organ: Mix of ovaries of varying immature stages
from 8-week-old plants; Vector: pBluescript II from Lambda
Zap II; Site_1: XhoI; Site_2: EcoRI; The library was made
from poly-A RNA in the cloning vector lambda Zap II.
Clones to be sequenced were prepared by mass excision."

BASE COUNT

129 a 172 c 191 g 99 t

ORIGIN

Query Match 3.4%; Score 78.6; DB 12; Length 591;
Best Local Similarity 57.1%; Pred. No. 8.7e-06;
Matches 172; Conservative 0; Mismatches 114; Indels 15; Gaps 1;
QY 396 GACCGGCTTCAAGGCTTCACTACCGGATCGCGAGCGCGCCGCGCCGCTTCGGGT 455
DB 219 GAACACCTTCCAGGGTTGGATTTCGTGCTAGCTAGGAGTACGGGATTAAAT 278
QY 456 GATTGTGCTCTCACCACCACTGGAAGAATTTGGGGGAATGATCAATAC----- 507
DB 279 GACTACTGAGGCTTTGTGAACAATTAACGATAGCTTTGGAGGAGGAGTACGTGCAGTG 338
QY 508 -----GACAAGTGGTACGGCTTCCTTACCAGCAAACTTCTACACGACCCCGGAC 560
DB 339 GCGGAGGAGCAGGGGAGGCCATTTCGATCTGACGATGAGTCTTCTACTAACCTGTTC 398
QY 561 CCAGCGGCTTACAAGAATTTGGGTCAATCATCTACTGAACCGGCTCAACAGCATTACCGG 620
DB 399 CAAGGGCTCTACAAGAACCATATCAAGGCGCTTCTCAGGAGGTGAACACGATCACCGG 458
QY 621 CGTGAGCTACAAGAACCATCAACGATCTTTCGTTGGAACTTCCCAATGACGCGCGCTG 680
DB 459 CGTGGCTACAAGGAGCAGCCCAACGATCATGGCGTGGGAGCTGATGAACGAGCGCGCTG 518
QY 681 C 681
DB 519 C 519

RESULT 7

BJ248065

LOCUS

BJ248065 655 bp mRNA linear EST 05-APR-2002
DEFINITION aestivum cDNA clone whf4122 5', mRNA sequence.

ACCESSION

BJ248065

VERSION

BJ248065.1

KEYWORDS

EST.

SOURCE

bread wheat.

ORGANISM

Triticum aestivum

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooidae

; Triticeae; Triticum.

1 (bases 1 to 655)

Ogihara,Y. and Murai,K.

Expressed genes in Triticum aestivum

Unpublished (2002)

Contact: Tadashi Shin-i

Center For Genetic Resource Information

National Institute of Genetics

1111 Yata, Mishima, Shizuoka 411-8540, Japan

Tel: 81-559-81-6856

Fax: 81-559-81-6855

Email: tshini@genes.nig.ac.jp.

Email: mmpratt@uga.edu

Sequences have been trimmed to exclude polyA, vector and regions
below Phred quality 16. The threshold for highest quality sequence
is 20.

Seq primer: JEN REV

High quality sequence stop: 514

POLYA=NO.


```

FEATURES
  source      Location/Qualifiers
1. .655
/organism="Triticum aestivum"
/cultivar="Chinese Spring"
/db_xref="taxon:4565"
/clone="whf4122"
/clone_lib="Y. Ogihara unpublished cDNA library, whf1"
/tissue_type="Spike at flowering date"
/dev_stage="Feekes' scale 10.5.1"
/Note="Vector: Lambda Uni-ZAP XR, excised phagemid;
Site_1: EcoRI; Site_2: XhoI; Plants were grown under
hydroponic conditions at UC Davis, salt stressed for 12
hours, and for 7 days, then dissected and frozen (Akhunov
in J. Dvorak Lab). Total RNA was prepared from sheath
tissue, equal quantities of RNA were pooled from the two
samples, polyA was purified from the pooled RNA, a cDNA
library was made, and the cDNA clones were in vivo
excised to give pBluescript phagemids in the TJ Close lab
at the University of California, Riverside (Akhunov, Chin
, Choi, Close, Fenton, Klianian, Otto, Simons, Zhang).
Plasmid DNA preparations and DNA sequencing were
performed in the OD Anderson lab (all other authors)."
```

BASE COUNT 116 a 214 c 227 g 98 t

ORIGIN

Query Match 3.4%; Score 78.6; DB 13; Length 655;
Best Local Similarity 52.3%; Pred. No. 8.9e-06;
Matches 242; Conservative 0; Mismatches 194; Indels 27; Caps 2;

```

QY 513 GTGTCAGCGCTTCCTTACCACGACAACTTCTACACGACCGCCGCGACCGACGCGGTA 572
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 6 GGGCAGGCGCTCGGATCTGAAGATGACTTCTTCCCAACTCCGTCGTAAGGGCTTCTA 65
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 573 CAAGAATTGGGTCAATCATCTACTGAACCGGCTCAACAGCATTTACCGCGCTGACGTACAA 632
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 66 CAAGAACCATTGTCAAGACCGCTGCTGACGAGGCTGAACACTGTGACTGGGTGGGTACAA 125
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 633 GAACGATCCACCATCTTTGCTTGGGAATTTGCCAATGACCGCGCTGCGTAGGAAGCGG 692
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 126 GAACGACCGGACGATCTTGGCGTGGGAGCTGATGAACGACCGCGGTGGCA----- 176
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 693 CACATTACCAACCTCGGACGCTGCACTCAGCGGACCATTTGTCAACTGGGTGATCAAAAT 752
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 177 -----GTCCGACCTCTCCGGCGGACCATCCATCGTGGATCAGGAGAT 221
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 753 GTCGCGGTACGTCAAAAGCATAGACCTTAACATATATGTTCTCGGTGGCGACAAAGGTT 812
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 222 GCGCGGCCAGCTCAAGTCCATCAGCGCAACCATGCTGTGAGGCGCGGCTGGAGGCTT 281
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 813 CTACATTGGGTCAACGAGGGAACGGCTGGCCATAC---AACGACCGCTCCACGCGGT 869
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 282 CTACGCGCGGTCTGCTCCCGCGCGCTCCGTTGAACCCGCGCGCGCGGAGTGGGCAC 341
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 870 CGACAACAATGTCTTCTCCGTCTAAGAACAATTGATTTGGCACGATATCACCTGTACCC 929
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 342 CGATTTTCATCGCAACAACAGGTCCCGCGCATCGACTTCGCCACCGTGCATCGTACCC 401
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 930 GAATTACTGGGCGCACACCGGACGTGGGAACGCAATGGATC 972
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 402 GGACCAAGTGGCTGTCGAGTCCGACGACGAGGCGGACGAGC 444
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 8
AV752548/c
LOCUS
DEFINITION NPd Homo sapiens cDNA clone NPDAZH01 5', mRNA sequence.
ACCESSION AV752548
VERSION AV752548.1 GI:10910396
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

```

```

REFERENCE
AUTHORS Song,H., Peng,Y., Gu,Y., Yang,Y., Gao,G., Xiao,H., Xu,X., Li,N.,
Qian,B., Liu,F., Qu,J., Gao,X., Cheng,Z., Xu,Z., Zeng,L., Xu,S., Gu
W., Tu,Y., Jia,J., Fu,G., Ren,S., Zhong,M., Lu,G., Ye,M., Zhang,Q.,
Han,Z., Chen,Z., Hu,R. and Chen,J.
TITLE Homo sapiens NPd library cDNA clones
JOURNAL Unpublished (2000)
COMMENT Shanghai Institute of Endocrinology, Rui-Jin Hospital
197 Rui-Jin II Road, Shanghai 200025, P. R. China
Tel: 86-21-64370045(ex.66332)
Fax: 86-21-64743206
Email: mbsliems.scn.sh.cn
This clone is available at Shanghai Hematology Institute in
Shanghai.
Chinese National Human Genome Center at Shanghai
351 Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong.
FEATURES
  source      Location/Qualifiers
1. .1315
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="NPDAZH01"
/clone_lib="NPd"
/tissue_type="pituitary"
/dev_stage="Adult"
/lab_host="SOLR"
/Note="Vector: pBluescript sk(-); Site_1: EcoRI; Site_2:
XhoI"
BASE COUNT 408 a 403 c 224 g 280 t
ORIGIN
Query Match 3.3%; Score 76.4; DB 10; Length 1315;
Best Local Similarity 66.3%; Pred. No. 2.9e-05;
Matches 110; Conservative 0; Mismatches 56; Indels 0; Caps 0;

QY 1203 CGCAGGAGCGCTCCCAATCAGTACCGGACATCCGCTCCGCGGAGCTCGAG 1262
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 192 CGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTC 133
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1263 TCCATCTCTCCGCGCTCCGTCGCGCTCGTCGTCGCGGCTCCGTCGCGGCTCCGTC 1322
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 132 GTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTC 73
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1323 GCGCTCTTCGTCGCGGAGCGCTCCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 1368
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 72 GTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTC 27
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 9
CNS04AAV 788 bp DNA linear GSS 21-MAY-2000
LOCUS Tetraodon nigroviridis genome survey sequence PUC-ori and of clone
DEFINITION 095x23 of library G from Tetraodon nigroviridis, genomic survey
sequence.
ACCESSION AL281632
VERSION AL281632.1 GI:8019957
KEYWORDS GSS: genome survey sequence.
SOURCE Tetraodon nigroviridis.
ORGANISM Tetraodon nigroviridis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
Tetraodontidae; Tetraodon.
1 (bases 1 to 788)
Roest-Crolius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C.,
Bernot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F.,
Saurin,W. and Weissenbach,J.
Human gene number estimate provided by genome wide analysis using
Tetraodon nigroviridis DNA sequence
Unpublished
REFERENCE
AUTHORS Roest-Crolius,H., Jaillon,O., Dasilva,C., Fizames,C., Fisher,C.,
Bouneau,L., Billault,A., Quetier,F., Saurin,W., Bernot,A. and

```

TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

Weissenbach, J.
Characterization and repeat analysis of the compact genome of the
freshwater pufferfish Tetraodon nigroviridis
Unpublished
3 (bases 1 to 788)
Genoscope.
Direct Submission
Submitted (12-APR-2000)
This sequence is a single read and was generated as part of a large
scale clone-end sequencing project of the Tetraodon nigroviridis
genome. For more information, please take a look at
http://www.genoscope.cns.fr/Tetraodon.

FEATURES
source
1..788
Location/Qualifiers
/organism="Tetraodon nigroviridis"
/db_xref="taxon:99883"
/clone_lib="G"
/note="Genoscope sequence ID : C08G095AF12SP1-end ;
PUC-ori"

BASE COUNT 192 a 235 c 213 g 131 t 17 others
ORIGIN

Query Match 3.3% Score 76.2; DB 17; Length 788;
Best Local Similarity 66.0%; Pred. No. 2.8e-05;
Matches 97; Conservative 8; Mismatches 42; Indels 0; Gaps 0;

QY 1809 TGAACCTCCGGCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAG 1868
::: : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : |||||
KBAYTBCCGGCGCTAGCCCTACKCYCAGTCTAGGCCCGCGCCCTACCCCTAGACCCAG 618
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1869 CCGAGCCGACCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAG 1928
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 619 CCCAGCCCTAGCCCTAACCTAGCCCGCCCTAGCCCTAGCCCTAGCCCTAGCC 678
QY 1929 CCCTACCCCTCCCGTCGCGGAGCC 1955
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 679 CCCTACCCCTAACCTAGCCCTAGCCCTAGCC 705

RESULT 10
AA720369
LOCUS
DEFINITION
33562 Lambda-PRL2 Arabidopsis thaliana cDNA clone 171N14T7, mRNA
sequence.

ACCESSION
AA720369
VERSION
AA720369.1 GI:2733979
KEYWORDS
EST.
SOURCE
thale cress.

ORGANISM
Arabidopsis thaliana

REFERENCE
AUTHORS
Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsi.
1 (bases 1 to 236)
Newman, T., deRuijn, F.J., Green, P., Keegstra, K., Kende, H., McIntosh
L., Ohlroge, J., Paik, N., Somerville, S., Thomas, M., Metzger
J., and Somerville, C.
Genotype: a summary of methods for accessing results from
a large scale parallel sequencing of anonymous Arabidopsis cDNA clones
Plant Physiol. 106, 1241-1255 (1994)

MSU Plant Research Laboratory
Michigan State University
MSU-DOE-PRL, Michigan State University, Plant Biology Bldg., E.
Lansing, MI

Tel: 517-353-0854
Fax: 517-353-9168
Email: 22313tcn@lm.cl.msu.edu
Seq primer: 17 dye primer.
Location/Qualifiers
1..236
/organism="Arabidopsis thaliana"

/strain="var columbia"
/db_xref="taxon:3702"
/clone_lib="171N14T7"
/note="Vector: Lambda Zip-Lox; Site_1: Sal; Site_2: NotI;
Lambda PRL2 is a cDNA library derived from equal
quantities of 4 pools of mRNA. The mRNA sources were 1) 7
day germinated etiolated seedlings; 2) tissue culture
grown roots; 3) stayed plants half with 24 hour light
cycle, half on 16 hr light, 8 hour dark- rosettes; 4)
same plants as 3 but aerial tissue (stems, flowers and
siliques). The vector is BRL's lambda Zip-Lox. The cDNA
inserts were directionally cloned with Sal-Not arms using
oligo dT primed cDNA."

BASE COUNT 31 a 79 c 46 g 71 t 9 others
ORIGIN

Query Match 3.3% Score 76; DB 9; Length 236;
Best Local Similarity 74.6%; Pred. No. 2.4e-05;
Matches 94; Conservative 0; Mismatches 32; Indels 0; Gaps 0;

QY 1246 CCGTCGCGGAGTCGAGTCATCTCGTCGCGGTCGTCGCGGTCGTCGCGGTCGTCG 1305
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2 CCGTCGCGGTCGCGTCGCTCCGTCGCTCCGTCGCTCCGTCGCTCCGTCGCTCCGTC 61

QY 1306 CCGTCGCGGTCGCTCCGTCGCGGTCGTCGCGGAGCCGTCGTCGTCGTCGTCG 1365
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 62 CCGTCGCGGTCGCTCCGTCGCTCCGTCGCTCCGTCGCTCCGTCGCTCCGTCGCTCC 121

QY 1366 TCGGGT 1371
|||
Db 122 CCGNAT 127

RESULT 11
AZ871772/c
LOCUS
DEFINITION
586 bp DNA linear GSS 21-FEB-2001
2M0184M17R Mouse 10kb plasmid UUGC1M library Mus musculus genomic
clone UUGC2M0184M17 R, DNA sequence.

ACCESSION
AZ871772
VERSION
AZ871772.1 GI:13078306
KEYWORDS
GSS.
SOURCE
house mouse.
ORGANISM
Mus musculus

REFERENCE
AUTHORS
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 586)
Dunn, D., Aoyagi, A., Barber, M., Boacorn, T., Duval, H., Hamil, C.,
Islam, H., Longacre, S., Mahmoud, M., Meenen, E., Pedersen, T., Reilly
M., Rose, M., Rose, R., Stokes, R., Tingey, A., von Niederhausern, A.
and Wright, D., Weiss, R.

TITLE
JOURNAL
COMMENT
Mouse whole genome scaffolding with paired end reads from 10kb
plasmid inserts
Unpublished (2000)
Contact: Robert B. Weiss
University of Utah Genome Center
University of Utah
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLIC, UT
84112, USA
Tel: 801 585 5606
Fax: 801 585 7177
Email: ddunn@genetics.utah.edu
Insert Length: 10000 Std Error: 0.00
Plate: 0184 row: M column: 17
Seq primer: CACACAGGAACAGCTATGACC
Class: plasmid ends
High quality sequence stop: 586.
Location/Qualifiers
1..586
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UUGC2M0184M17"

[illegible]

Inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adapted mouse DNA was annealed to adapted vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

BASE COUNT 110 a 151 c 183 g 172 t
ORIGIN

Query Match 3.2%; Score 72.4; DB 17; Length 616;
Best Local Similarity 77.2%; Pred. No. 0.00016;
Matches 88; Conservative 0; Mismatches 26; Indels 0; Gaps 0;
Qy 1824 CAGCCCCAGCCCCACACCCAGCCCCAGCCCCAGCCCCGTCCTCCCGAGCCCCAGCCCCGAGCCCCGAGCCCC 1883
Db 121 CAGCCCCAGCCCCAGCCCCAGCCCCAGCCCCAGCCCCAGCCCCAGCCCCAGCCCCAGCCCCAGCCCCAGCCCC 62
Qy 1884 AAGCCCCAGCTCTCCCATCTCCCGTCCCGAGCCCCAGCCCCAGCCCCAGCCCCAGCCCCAGCCCCAGCCCC 1937
Db 61 CAGCCCCAGCCCCAGCCCCAGCCCCAGCCCCAGCCCCAGCCCCAGCCCCAGCCCCAGCCCCAGCCCCAGCCCC 8

Search completed: November 13, 2002, 10:42:52
Job time : 2157 secs

Db 2078 CCAGGCGGTCCGCGAGTCGCGAGCCGACGCTACTCCGACGCGACGACGCGCGAGC 2137
QY 1912 CCAGGCGCGAGCCGAGCCGAGCCGCTCCCGTCGCGAGCGCGTCGCGCGTCCGCGAGT 1971
Db 2138 CCAACGCTGACCCCTACTGCTAGCCGACCGCCGACGCGCAAGCCGCGTCCCGG-- 2195
QY 1972 GTCTGCTGCTGCGGTGTCGCGGCGCGACGCTATGCTGTAATAGTATTGGGGTCT 2031
Db 2196 -CGGACGCTCCGAGCGCGCTGACCGCGAGTTACAGGTCAACAGGATTGGGCAAT 2254
QY 2032 GGGTTTACGCGGAGCGGTGAGCGGTGACGAATACCGGGAGCGCGGCGAGCGGGTGGAGC 2091
Db 2255 GGGTTTACGCGTAAAGCGGTGGCGGTGACAAATTCGGGATCCGTCGCGACCAAGACATGAGC 2314
QY 2092 GTGCGGTGGTCTGTTGGTGGGAATCAGACGCTCAGCAACTACTGGAACACTGCGTTGACC 2151
Db 2315 GTGAGTTGGACATTCGGCGGAATCAGACGATTACCAATTCGTTGGAATGAGCGGTCACG 2374
QY 2152 CAATCAGGTGCATCGGTGACGGCGAGCAACCTGAGCTTACAAACGCTGATCCAAACCGGT 2211
Db 2375 CAGAAGCGTCAGTCGGTAAACGCTCGGAATATGAGTTATAACAACGTTGATTCAGCCTGGT 2434
QY 2212 CAGTCGACCACTTCGGATTCAACGGAATTTACTCAGGAACAACACCGCGACCTACACTC 2271
Db 2435 CAGAACACCACTTCGGATTTCAGGCGAGCTATACCGGAAGCAACCGCGGACCGACAGTC 2494
QY 2272 ACCTGACCGCTAGTTGA 2289
Db 2495 GCCTGCCACCAAGTTAA 2512

RESULT 2

US-09-136-574A-1

; Sequence 1, Application US/09136574A
; Patent No. 6294366

GENERAL INFORMATION:

; APPLICANT: Farrington, Graham K.

; Anderson, Paige

; Gibbs, Moreland

; Berquist, Peter

; Daniels, Roy

; Morgan, Hugh W.

; Williams, Diane P.

; TITLE OF INVENTION: Compositions and Methods for

; Treating Cellulose Containing

; Cellulase Enzyme Compositions

; NUMBER OF SEQUENCES: 49

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Howson and Howson

; STREET: Spring House Corporate Center, P.O. Box 457

; CITY: Spring House

; STATE: PA

; COUNTRY: USA

; ZIP: 19477

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Diskette

; COMPUTER: IBM Compatible

; OPERATING SYSTEM: DOS

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/09/136,574A

; FILING DATE: 19-Aug-1998

; CLASSIFICATION: <unknown>

; PRIORITY INFORMATION:

; APPLICATION NUMBER: US 08/932,571

; FILING DATE: September 19, 1997

; ATTORNEY/AGENT INFORMATION:

; Name: Bak, Mary E.

; REGISTRATION NUMBER: 31,215

; REFERENCE/DOCKET NUMBER: 1997US001/CIP

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 215-540-9200

; TELEFAX: 215-540-5818

;

; TELEX: <Unknown>
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 11707 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; SEQUENCE DESCRIPTION: SEQ ID NO: 1:
; US-09-136-574A-1

Query Match 5.8%; Score 133.6; DB 4; Length 11707;
Best Local Similarity 52.4%; Pred. No. 4.2e-20;
Matches 395; Conservative 0; Mismatches 344; Indels 15; Gaps 4;

QY 1220 CAATCAGTACCGGCGACATCGCCTCCGCGGTGCGGAGCTCGAGTCCATCTCTGTCGCGGT 1279
Db 3883 CAACACACACCTGCACCGACATCAAGCCCGGACAGCCGCAACAGTACAGCAACGCGGA 3942
QY 1280 CTCGTCGCGGTCTCCGTCGCGGTCTCCGTCGCGGTCTCCGTCGCGGTCTCTTCTGTCGCGGA 1339
Db 3943 CTCACGCGCGCGCGACAGTACGAGTTACTGTGACTCCGACACCGACACACACCGA 4002
QY 1340 GCCGTCCTCGTCGTCGTCGCGGTGTCGG---GTGGGGTGAAGGTCCAGTACAGCAACA 1396
Db 4003 CGCGCACACGACAGGACACCTGGCACGGGAAGTGGTTGAAGGTACTATACAAGAACA 4062
QY 1397 ATGATTCCGCGCGCGGTGATACACGATCAACCGGGTCTCCAGTTCGTTGATACAGGGGT 1456
Db 4063 ATGAGCAAGTCGCGAGCACAAAGTTCTATAAGCGCGTGGTTAAGATAGTGAATGGAGGCA 4122
QY 1457 CGTCGTCGTTGATTTGTCGACGCTGACGCTGCGGTACTGTTCCACCGCGGATCGTGGT 1516
Db 4123 GCAGCAGTGTGATCTTTAGCAGGCTTAAGATAAGATACTGTTACACAGTGGATGCTG--- 4179
QY 1517 CGTCGACACTGTTGTACAACTGTGACTGGCGGGGATGGGTGGGAATATCCGCCCTT 1576
Db 4180 ACAAGCCACAGAGTGGCGGTATGTACTGGGCACAGATAGGGGCAAGCAATGTCACATTCA 4239
QY 1577 CGTTCCGCTCGGTGAACCGCGGACGCGGACGCGGACACTACCTGCAGTTGTCGTTCA 1636
Db 4240 ATTTGTGAAGCTGAGCAGCGGAGTGAAGTGGCGGATTTATTCTTGAGGTAGGATTTA 4299
QY 1637 CTGCTGGAAC-----GTGGCGCGCTGGTGGGTGCGAGCGGTGAGATTCAAAACCGGTTGA 1690
Db 4300 CGAGTGGAGCTGGGACGTTCCAGCCTCGTTAAGGACACAGGGGATATACAGGTAAGTTTA 4359
QY 1691 ATAGAGTACTGTCGAACCTTTGATGAGACCAATGACTACTGTTGAGGACGACGACCG 1750
Db 4360 ACAAGATGACTGGAGCAATTTACAATCAGGCACGACGACTGCTATGTTGTCGAGAGCATGA 4419
QY 1751 CCTTCCAGGATTTGGA---CGAAGTGAAGTGTATGTCATGGCGGCTGTTGGTGGGGA 1807
Db 4420 CGAATTATGAGAGAAATGCGAAGTAAAGTGTATGTTAGTGGTGTCTGGTATGGGGG 4479
QY 1808 CTGAACGCTCCGCGACACAGCCCGACAGCCCGACAGCCCGACAGCCCGACAGCCCGTCCCGGA 1867
Db 4480 AGGAGCGCGCGGACGACACCTGACCGGACAGCAGCAGCAACCAACCAACGCAACGCGCA 4539
QY 1868 GCCCGAGCG 1927
Db 4540 CAGCAACAGCAACACCG 4599
QY 1928 GCCTACGCGCGTCCGCGTCCGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1961
Db 4600 CACCAACCCCAATACCAACACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 4633

RESULT 3

US-09-136-574A-2

; Sequence 2, Application US/09136574A

; Patent No. 6294366

; GENERAL INFORMATION:

; APPLICANT: Farrington, Graham K.

Anderson, Paige
Gibbs, Moreland
Bergquist, Peter
Daniels, Roy
Morgan, Hugh W.
Williams, Diane P.
TITLE OF INVENTION: Compositions and Methods for Treating Cellulose Containing Fabrics Using Truncated Cellulase Enzyme Compositions
NUMBER OF SEQUENCES: 49
CORRESPONDENCE ADDRESS:
ADDRESSEE: Howson and Howson
STREET: Spring House Corporate Center, P.O. Box 457
CITY: Spring House
STATE: PA
COUNTRY: USA
ZIP: 19477
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/136,574A
FILING DATE: 19-Aug-1998
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/932,571
FILING DATE: September 19, 1997
ATTORNEY/AGENT INFORMATION:
NAME: Bak, Mary E.
REGISTRATION NUMBER: 31,215
REFERENCE/DOCKET NUMBER: 1997US001/CIP
TELECOMMUNICATION INFORMATION:
TELEPHONE: 215-540-9200
TELEFAX: 215-540-5818
TELEX: <Unknown>
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 6416 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
SEQUENCE DESCRIPTION: SEQ ID NO: 2:
US-09-136-574A-2
Query Match 5.6%; Score 128.4; DB 4; Length 6416;
Best Local Similarity 52.6%; Pred. No. 4.9e-19;
Matches 379; Conservative 0; Mismatches 326; Indels 15; Gaps 4;
QY 1243 CCGCCGTCGCGGAGTCGAGTCCTCTCGTCGCGCGTCTCGTCGCGCGTCTCGTCGCGGCG 1302
DB 2533 CCGGGTCGCGCAACATCCACACCTGCAACCAATCAACACCAACACCGGACCCCA 2592
QY 1303 TCCTCGGTCGCGGTCCTCGTCGCGGAGCCGTCCTCGTCGTCGTCGTCGTCGCGG 1362
DB 2593 ACACCAACAGTGACACGACGCGGACGCTCTACACCGACGACGCGGTCACCT 2652
QY 1363 GTCTCGG---GTGCGGTGAAGTGCACTACAGCAACAAATGTCGCGCGCGGTGATAAC 1419
DB 2653 GGTACGGGAAGTGTGTGAAGTACTGTACAGCAACAAATGAGACAGTGGGACACAGGT 2712
QY 1420 CAGATCAACCGGTCCTCCAGTTGGTGAATACAGGGGTCTGTCGTCGTCGTCGTCGACG 1479
DB 2713 TCTATAAGCCGCTGGTTAAGATAGTGAATGGAGGACGACGAGTGTGATCTTAGCAGG 2772
QY 1480 GTACCGGTGCGGTACTGGTTACCGCGGATGGTGGTCTGTCGACACTGGTGTACAACGT 1539
DB 2773 GTTAAGATAGTACTGGTACACAGTGGATGGTG---ACAAGCCACAGAGTGGCGGTATGT 2829
QY 1540 GACTGGCGCGGATGGGTGTGGGAATATCCCGCGCTCTCGGCTCGGTGGAACCGCGCG 1599
DB 2830 GACTGGGCACATAGGGGCAAGCAATGTGACATTTCAATTTTGTGAAGCTTACCGCGGA 2889

QY 1600 ACGCCGACGGCGGACACCTACCTGCAGTTGCTGTTCACTGCTGGAAC-----GTTGGCC 1653
DB 2890 GTGAGTGGAGCGGATATTACCTGGAGCTAGGATTTACGAGTGGAGCTGGCAGTTGCAG 2949
QY 1654 GCTGTTGGTTCAGCGGTGAGATTCAAAACCGGGTGAATAAGAGTGACTGTCGAACTTT 1713
DB 2950 CTTGTAAGGACACAGGGGATATACAGGTAAGGTTTAAACAAGAATGACTGGAGCAATTAC 3009
QY 1714 GATCAGACCAATGACTACTCTGATGGGACGAACACCGCTTCCAGGATTGA---CGAAG 1770
DB 3010 AATCAGGACGAGACTGGTCTATGGTTGCAGAGCATGACGAATTTGGAGAGAATCCGAAG 3069
QY 1771 GTGACGTTGATGTCAATGGCCGCTGCTGTTGGGGAGTGAACCGTCCGGACACCGCC 1830
DB 3070 GTGACGCTGTATGTAGATGGTGTCTGCTATGGGGCAGGAGCCGGAGGAGGACACCT 3129
QY 1831 AGCCCCACACCCAGCCCGCCAGCCCAACCCGTCCTCCGGAGCCCGAGCCCGACCCAGCC 1890
DB 3130 GCACCGACAGCAGCAGCAACACCAACCGCAACTCCGACAGCAACCCCAACCTACACCT 3189
QY 1891 AGCTCTCCCGCATCCCGTCCCGGAGCCCGCCAGCCCGCTAGCGCTCCCGCTCGCGG 1950
DB 3190 ACACCGACCCGACACCGACAGTGTGTCACAGCCGACCCGCGACCGCGCATCACCG 3249
RESULT 4
US-09-134-078-57
Sequence 57, Application US/09134078
Patent No. 638844
GENERAL INFORMATION:
APPLICANT: Bylina, Edward J.
TITLE OF INVENTION: GLYCOSIDASE ENZYMES
NUMBER OF SEQUENCES: 72
CORRESPONDENCE ADDRESS:
ADDRESSEE: Gray Cary Ware & Freidenrich LLP
STREET: 4365 Executive Drive, Suite 1600
CITY: San Diego
STATE: CA
COUNTRY: USA
ZIP: 92121
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: Windows95
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/134,078
FILING DATE: 13-AUG-1998
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/949,026
FILING DATE: 10-OCT-1997
APPLICATION NUMBER: 60/056,916
FILING DATE: 06-DEC-1996
ATTORNEY/AGENT INFORMATION:
NAME: Halle, Lisa A.
REGISTRATION NUMBER: 38,347
REFERENCE/DOCKET NUMBER: 09010/024002
TELECOMMUNICATION INFORMATION:
TELEPHONE: 858/677-1456
TELEFAX: 858/677-1465
INFORMATION FOR SEQ ID NO: 57:
SEQUENCE CHARACTERISTICS:
LENGTH: 1992 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: Genomic DNA
FEATURE:
NAME/KEY: Coding Sequence
LOCATION: 1...1989
US-09-134-078-57

Query Match		5.5%	Score 126.4;	DB 4;	Length 1992;
Best Local Similarity		49.3%	Pred. No. 9.8e-19;		
Matches 513;		Conservative	0;	Mismatches 476;	Indels 51; Gaps 5;
Qy	121	TTCTGTCACCGCATCCGGCGGTCACTTCGTCTGAACCGCCTTCCTCTATCGTTACGGGGGA	180		
Db	49	TTCTGTAAGTGGAAACCGGAAATTCGCTCTGAACGAAAGAAATTCAGATTCATTCGA	108		
Qy	181	ACGAACAATATTACCTCAGCTATCATCGCACCGCCAGCTCGATGACGTGTTCCGCCAAG	240		
Db	109	AGCAACAATCTACATGCATACAAAGCAGCAAGCAATGATAGACAGTGTTCGGAGAGT	168		
Qy	241	GCTCAACGGATGAATCTTTCTGTCTACCGGACCTGGGTTTCATCGACATCGGCTCTCTT	300		
Db	169	GCCAGACATATGATATAAAGTCTCTCAGAACTCTGGGTTTCTCTCGACGGGAGAGTTAC	228		
Qy	301	GACGGCTCCGTGCCACAATCGATGGCAACAAGAACGGCTTCTACTTTTCAGTACTGGAC	360		
Db	229	TGCAGACAGAAGACATCATCATCTGAGCCCGGTGTTTCGGGGTGCCAGAGGA	288		
Qy	361	CCGTCCGACCGCGCTCCGGGTACAACGCGCGCCGCTTGAAGGCTTGAAGTAC	420		
Db	289	ATATCGAACGCC-----CAGAGCGGTTTCGAAGACTCGACTAC	327		
Qy	421	GCGATCCGAGCGGGCGCGCAGCGCTTCGGGTGATGTGCTCTCAACCAAGACTGG	480		
Db	328	ACAGTTGCGAAGCGGAAGAACTCGGTATAAACTTCTCATTTGTTTGAACAACCTGG	387		
Qy	481	AAGAAATTCGGGAATGGATCAATACGACAGTGTACCGCTTCCTTACCAGCACAAC	540		
Db	388	GACGATTCGGTGAATGAACAGTACGTAGGTGGTGGGAGACCCATCAGACGAT	447		
Qy	541	TTCTACACCCACCCCGGACCGGCTGAGAGCGGTACAAGATTTGGGTCAATCTACTGAAC	600		
Db	448	TTCTACAGATGAGAAGATCAAGAAGAGCTACAAAAGTACGTCCTTCTCTGTAAAC	507		
Qy	601	CGGTCAACAGCATTACCGGGTGACGTACAGAAGCAATCCAAGCATCTTTGCTTGGAA	660		
Db	508	CATGTCAATACCTACACCGGAGTTCTTACAGGAGAGCGCCACCATCATGCGCTGGAG	567		
Qy	661	CTTGCCAAATGAGCGCGCTCGTAGGAAGCGGCACATTACCAACCTCGGCACCTGCAC	720		
Db	568	CTTGCAACCAACCGCGCTGTAGACGAGCAATTCGGGGACACGCTCTGTGAGTGGGTG	627		
Qy	721	CAGCGACCAATTTGTAACCTGGGTGATCAAAATGTGCGGTACGTCAAAAGCATAGACCT	780		
Db	628	AAGGAGA-----TGAGTCTCTACATAAAGAGTCTGCATCCCAACCACTCGTGGCTGTG	681		
Qy	781	AACCATATGCTCTCGTTCGGCGGACGAGGTTCTACATTGGGTCAACGCAAGGAGCGGC	840		
Db	682	GGGACCAAGGATTTCTCAGCAACTACGAAGGATTTCAAACTTTACGGTGGAGAGCCGAG	741		
Qy	841	TGGCCATACAAACACCGCTCCGACGGGTCCGACAACATGCTCTTCCTCGGTGTCAAGAAC	900		
Db	742	TGGGCTTACAAACCGGTGTC---CGGTGTGACTGGAAGAGTCTCTTCGATAGAGAGC	798		
Qy	901	ATTGACTTTGGCAGTATACCTGTATCCCGAATTTACTGGGG-----CCAGAAC	948		
Db	799	GTGACTTCGCCACGTTCACCTCTATCCGTCCCACTCGGGGTGTCACTCCAGACAATAT	858		
Qy	949	GCGGACTGGGGAACGCAATGATCAAGGATCATATTCGGAATCCGCGCAGCGATCGGCAAG	1008		
Db	859	GCCAGTGGGGGAGGTGGATAGAAGACCAATAAAGATTCGCAAAAGAGATCGGAAAA	918		
Qy	1009	CCGACCAATTCGGAAGAAATTCGG-----CTGGCAGACACCGGACCGGATTCGGTC	1059		
Db	919	CCCGTTGTTCTGGAAGAATATGNAATTCCAAGAGTCCGCCAGTATACAAACGGCCATC	978		
Qy	1060	TATCAGACGTGGACCAAGCTGTGCGGTACGAACGGGTGAACAGCTGGAACTTCTGGATG	1119		
Db	979	TACAGACTCTGGGAACGATCTGGTCTACGATCTCGGTGGAGATGGAGCGATGTTCTGGATG	1038		
Qy	1120	CTCGCTGGGAATCTCAACGG	1139		

Db	1039	CTCGCGGAATCGGGGAAGG	1058		
RESULT 5					
US-09-134-078-11					
; Sequence 11, Application US/09134078					
; Patent No. 6368844					
; GENERAL INFORMATION:					
; APPLICANT: Bylina, Edward J.					
; TITLE OF INVENTION: GLYCOSIDASE ENZYMES					
; NUMBER OF SEQUENCES: 72					
; CORRESPONDENCE ADDRESS:					
; ADDRESSEE: Gray Cary Ware & Freidenrich LLP					
; STREET: 4365 Executive Drive, Suite 1600					
; CITY: San Diego					
; STATE: CA					
; COUNTRY: USA					
; ZIP: 92121					
; COMPUTER READABLE FORM:					
; MEDIUM TYPE: Diskette					
; COMPUTER: IBM Compatible					
; OPERATING SYSTEM: Windows95					
; SOFTWARE: FastSeq for Windows Version 2.0					
; CURRENT APPLICATION DATA:					
; APPLICATION NUMBER: US/09/134.078					
; FILING DATE: 13-AUG-1998					
; CLASSIFICATION: 435					
; PRIOR APPLICATION DATA:					
; APPLICATION NUMBER: 08/949,026					
; FILING DATE: 10-OCT-1997					
; APPLICATION NUMBER: 60/056,916					
; FILING DATE: 06-DEC-1996					
; ATTORNEY/AGENT INFORMATION:					
; NAME: Halle, Lisa A.					
; REGISTRATION NUMBER: 38,347					
; REFERENCE/DOCKET NUMBER: 09010/024002					
; TELECOMMUNICATION INFORMATION:					
; TELEPHONE: 858/677-1456					
; TELEFAX: 858/677-1465					
; INFORMATION FOR SEQ ID NO: 11:					
; SEQUENCE CHARACTERISTICS:					
; LENGTH: 2043 base pairs					
; TYPE: nucleic acid					
; STRANDEDNESS: double					
; TOPOLOGY: linear					
; MOLECULE TYPE: Genomic DNA					
; FEATURE:					
; NAME/KEY: Coding Sequence					
; LOCATION: 1...2040					
US-09-134-078-11					
Query Match					
Best Local Similarity					
Matches 513; Conservative					
0; Mismatches 476; Indels 51; Gaps 5;					
Qy	121	TTCTGTACCGCATCCGGCGGTCACTTCGTCTGAACCGCCTTCCTCTATCGTTACGGCGGA	180		
Db	100	TTCTGTAAGTGGAAACCGGAAATTCGCTCTGAACGAAAGAAATTCAGATTCATTCGA	159		
Qy	181	ACGAACAATATTACCTCAGCTATCATCGCACCGCCAGCTCGATGACGTGTTGGCCAAG	240		
Db	160	AGCAACAATCTACATGCATACAAAGCAGCAAGCAATGATACACAGTGTTCGGAGAGT	219		
Qy	241	GCTCAACGGATGAATCTTTCTGTCTACCGGACCTGGGTTTCATCGACATCGGCTCTCTT	300		
Db	220	GCCAGACATATGGGTATAAAGTCTCTCAGAACTCTGGGGTTTCTCTCGACGGGAGAGTTAC	279		
Qy	301	GACGGCTCCGTGCCACAATCGATGGCAACAAGAACGGCTTCTACTTTTCAGTACTGGGAC	360		
Db	280	TGCAGACAGAAGAACACCTACATGCATCTCGAGCCCGGTGTTTTCCGGGGTGCCAGAGGA	339		
Qy	361	CCGTCCGACCGCGCTCCGGGTACAACGACGGCGCCAGCTTGAAGGCTTTGACTAC	420		

Db 340 ATATCGAACGCC-----CAGAGCGGTTTCGAAAGACTCGACTAC 378
QY 421 CGCATCGGACGGCGCGCGACGGCTTCGGGTGATTCGCTCCTCACCACGACTG 480
Db 379 ACAGTTGCCAAGCGAAGAACTCGGTATATAAATTCGTATTCCTTGTGAACAAC 438
QY 481 AAAGAAATTTGGGGGAATGGATCAATACGACAAAGTGTACGGCTTCCTTTACCACGACAAC 540
Db 439 CACGACTTCGGTGGAAATGAACCACTACGTGAGTGTGTTGGAGAACCCATCAGCAGCAT 498
QY 541 TTCTACACGACCCCGGACCGGAGCGGTACACAGAAATGGGTCAATCATCTACTGAAC 600
Db 499 TTCTACAGAGATGAGAAGATCAAGAAAGAGATACAAAAGTACGTCTCTTTCTCGTAAAC 558
QY 601 CGGGTCAACAGCATTTACCCGGGTGACGTACAAAGAACGATCCAAACGATCTTTGCTGGGAA 660
Db 559 CATGTCAATACCTACACGGGAGTCTCTTACAGGAAGAGCCCAACCATCATGGCTGGGAG 618
QY 661 CTTGCCAATGACCGCGGCTCGGTAGGAAGCGGCACATTAACCAACCTCGGGCAGTGCAT 720
Db 619 CTTGCAACGAACCGCGCTGTGAGCGGACAAATCGGGGAACACGCTCTGTGAGTGGTG 678
QY 721 CAGGCGACCATTTCAACTGGGTGCGATCAATCTCGCGCTAGCTCAAAAGCATAGACCCT 780
Db 679 AAGGAGA-----TGAGCTCTACATAAAGAGTCTGGATCCCAACCACTCGTGGCTGTG 732
QY 781 AACCATATGCTCGGTGCGGACGAAGGGTTCTACATTGGGTCAACGAGCGAAGCGGC 840
Db 733 GGGGACGAAGGATCTTTCAGCAACTACGAAGGATTCAAACCTTACGGTGGAGAAGCCGAG 792
QY 841 TGGCCATACAGACCGGCTCCGACGGGCTGCGACACAATGCTCTTCCTCGGTGTCAAGAAC 900
Db 793 TGGGCGCTACAACGGGTGGTFC---CGGTGTGACTGGAAGAAGCTCTTTTCGATAGAGACG 849
QY 901 ATTGACTTTGGCAGCTATCACTGTACCCGTAATTTACTGGG-----CCAGAAC 948
Db 850 GTGGACTTCGGCAGCTTCCACCTCTATCTCGTCCCACTGGGGGTGTCAGTCCAGAGAACTAT 909
QY 949 GGGGACTGGGGAACCGCAATGGATCAAGGATCATATTCGGAATCCCGACGCGATCGGCAAG 1008
Db 910 GCCCAGTGGGAGCGAAGTGGATAGAGACCAACATATAAGATCGCAAAACAGATCGGAANA 969
QY 1009 CCGACCATTTCTCAAGAATTCGG-----CTGGCAGACACCGGACCGCGATTCGGTC 1059
Db 970 CCGGTGTGTTCTGGAAGAATATGGAATTCGGAAGAGTGGCCCAAGTTAACAGAACGGCCATC 1029
QY 1060 TATCAGACTGGACCCAGACTCTGCGTACGAACGGTGAAGCAGLCTGGAACCTTCGGGATG 1119
Db 1030 TACAGACTCTGGAACGATCTGCTACGATCTCGGTGGAGATGGAGCGATGTTCTGGATG 1089
QY 1120 CTCGCTGGGAATGTCAACGG 1139
Db 1090 CTCGCGGGAATCGGGGAGG 1109

RESULT 6

US-08-525-697-1
: Sequence 1, Application US/08525697
: Patent No. 5795764
: GENERAL INFORMATION:

: APPLICANT: Christgau, Stephan
: APPLICANT: Andersen, Lene N
: APPLICANT: Kauppinen, Sakari
: APPLICANT: Heldt-Hansen, Hans P
: APPLICANT: Dalboe, Henrik
: TITLE OF INVENTION: AN ENZYME EXHIBITING MANNANASE ACTIVITY
: NUMBER OF SEQUENCES: 15
: CORRESPONDENCE ADDRESS:
: ADDRESS: No. 57957640 No. 5795764th America, Inc.
: STREET: 405 Lexington Avenue, 64th Floor
: CITY: New York
: STATE: New York

: COUNTRY: United States of America
: ZIP: 10174-6401
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patent In Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/525,697
: FILING DATE: 21-SEP-1995
: CLASSIFICATION: 435
: ATTORNEY/AGENT INFORMATION:
: NAME: Harrington, James J.
: REGISTRATION NUMBER: 38,711
: REFERENCE/DOCKET NUMBER: 4004.204-US
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 212-867-0123
: TELEFAX: 212-878-9655
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: DK 0486/93
: FILING DATE: 30-APR-1993
: CLASSIFICATION: 435
: INFORMATION FOR SEQ ID NO: 1:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1302 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: FEATURE:
: NAME/KEY: CDS
: LOCATION: 61..1192
: US-08-525-697-1

Query Match 5.2%; Score 118.6; DB 1; Length 1302;
Best Local Similarity 49.6%; Pred. No. 4.4e-17;
Matches 436; Conservative 0; Mismatches 389; Indels 54; Gaps 3;
QY 351 GTACTGGGACCGGCTCCGCGGTACACGACGAGCGGCGGCGGCTTGAAGG 410
Db 357 GTACAGCTGCAGCGACGGCACCTCGACCATCAACACGGCGCGGCTCCAGCG 416
QY 411 CTTTGAATACGGGATCGGAGCGCGCGGCGGCGGCTTGGGTGATTCGCTCTAC 470
Db 417 CCTGACTACGTGGTGCCTCGCGCGAGAGTACGGCGTCAAGCTGATCAACTCGT 476
QY 471 CAAGGACTGGAAGAATTTGGGGAATGGATCAATACGACAGTGGTACGGCTTCTTA 530
Db 477 CAACGAGTGGACCGGATCGGCGGCTACGAGGCTACGCGGCTACGCGGCGGCGG 536
QY 531 CCACGACAACTTCTACACGACGCGCGGCGGCGGCGGCTTGAAGGCTTGAAGTCA 590
Db 537 CCAGACGGACTTCTACACCAACACCGCGCTCCAGCGGCGGCTACACAGACTACATCAAGCG 596
QY 591 TCTACTGAACCGGCTCAACAGCATTTACCGGCGGTGAGCTACAGAGCATCCCAAGATCT 650
Db 597 GGTGCTCTCGCGGTACAGCAGCTCCCGCG-----CCATCTT 632
QY 651 TGTCTGGGAATTCGCAATGAGCGCGCTGCGTACGAGCGGCGGCGGCTTACCAACCTCGG 710
Db 633 CGCTGGGAGCTGGCGCAACGAGCGCGCTGCCAGGCGCTCGGATACCT----- 679
QY 711 CAGTGCATCTCAGGCGGACCATTTGTCAACTGGGTGATCAAAATGTCGGCGTACGCTCAAAG 770
Db 680 -----CGGTCTCTGATCAAACTGGATCTCGGACAGCTCCCAAGTATATCAAGTC 725
QY 771 CATAGACCTTAACCATATGCTCGGTGCGGCGGCGGAGGCTTCTACATTTGGGTCAACGCA 830
Db 726 GCTGAGCTCCAAGCACTGGTTCAGGATTCGCGATGAGGCTTCGGTCTCGAGCTCGACTC 785
QY 831 GGAAGCGGCTGGCGATACAGACCGCTCCGACGCGGCTCCGACAACTGCTCTCTCCG 890
Db 786 CGACGCGGCTACCCCTACACCTAC-----GGCGAGGATTTGAACCTTCCACCAAGAACCTGGG 842

Qy 891 TGTCAAGAACATTGACTTTGGCAGCTATCACCTGTACCCGAATTAATCTAGTGGGCGAGAACGC 950
Db 843 CATCTCGACCATCGACTTCGGCTACTCTGCATCTGTACCCCGATAGCTGGGCGACCTCTCTA 902
Qy 951 GGACTGGGAGCAAGTCAAGGATCATATTCCGAATGCCCGAGGATCGGCAAGCC 1010
Db 903 CGACTGGGGCAACGGCTGGATCAGCGGCCCGCCCGCCCTGCAAGCGGCTGGCGAAGCC 962
Qy 1011 CACCATTCCTCAAGAATTCGGCTGCGACACACCGGACCGGATTCGGTCTATCAGACCTG 1070
Db 963 GTGCTCTGGAAGACTACGAGTGAACCTCCAACCACTGTGCGCTCGAGAGCCCTGGCA 1022
Qy 1071 GACCCAGACTGTGGCTACGAAGGTGAAGCAGGTGGAACTTCTGGATGCTCGCTGGGAA 1130
Db 1023 GCAGACGGCGGCAACGCGAGCGGCGCATCTCCGGCGATTGTACTGGCAGTATGGCAACC 1082
Qy 1131 TGTCAAGCGCCAGCCATATCCGAACATATGACGGCTTCAACCTCTACTACCCAAAGTTCAAC 1190
Db 1083 GTTCAGCTGGGGCGGAGTCCCCGAGCATGAGGATGGGAACACCTTCTACTACAAACACAGCGACTT 1142
Qy 1191 AGCAGCCCTCTCGCCAGCGAGGCGCTCGCAATCAGTAC 1229
Db 1143 CACGTGCTGTGACGGATCATGTGGCGGCCATCAATCC 1181

RESULT 7

US-07-862-588B-1
; Sequence 1, Application US/07862588B
; Patent No. 5916796
; GENERAL INFORMATION:
; APPLICANT: Joergensen, Per Linea
; APPLICANT: Schlein, Martin
; APPLICANT: Hansen, Christian
; TITLE OF INVENTION: An Enzyme Exhibiting Cellulase Activity
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 59167960 No. 5916796disk of No. 5916796th America, Inc.
; STREET: 405 Lexington Avenue, 62nd floor
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10017
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/862.588B
; FILING DATE: 19920727
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: DK 164/90
; FILING DATE: 19-JAN-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/DK91/00013
; FILING DATE: 18-JAN-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Zelson, Steve T. / Lambiris, Elias J.
; REGISTRATION NUMBER: 30,335 / 33,728
; REFERENCE/DOCKET NUMBER: 3425.204-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212 867 0123
; TELEFAX: 212 867 0298
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2977 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: double
; TOPOLOGY: linear
; Molecule TYPE: DNA (genomic)
; ORIGINAL SOURCE:
; ORGANISM: Bacillus lautus

; STRAIN: NCIMB 40250
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 677..2776
; OTHER INFORMATION:
US-07-862-588B-1

Query Match 5.0%; Score 114.8; DB 2: Length 2977;
Best Local Similarity 57.1%; Pred. No. 3.6e-16;
Matches 250; Conservative 0; Mismatches 182; Indels 6; Gaps 2;

Qy 1382 TCCAGTACAGAACAAATGATTGGCGCCGGGTGATAACACAGATCAAAACCGGGTCTCCAGT 1441
Db 148 TGCATACAGAGCGCCGATACAAATGACGCGCAACACAGATCAAGCGGCTCTTCAACA 207
Qy 1442 TGGTGAATACGGGTGCTGCTGGGTGGATTTGTCACGGTGACGGTGCGGTACTGGTTCA 1501
Db 208 TCAAAACAAACGGTACTTCCGGCTGTGATTTAAGCACGCTCAAAATCCCGTACTACTTCA 267
Qy 1502 CCCGGGATGGGTGCTGCGACACTGGTGTACAACCTGTGACTGGCGGCGGATGGGTGTG 1561
Db 268 CCAAGGATGGTCTCGCGGGGTGAACGGCTGGATC---GACTGGGCGGAGCTCGGCGGCA 324
Qy 1562 GGAATATCCGCGCTCTGCGGCTCGGTGAACCCCGGCGGCGGAGACCTATC 1621
Db 325 GCAACATTGAGATCTCGTTTGGCAACCATACTGGCACGAATTCGGATACGTACGTGGAGC 384
Qy 1622 TCCAGTTGCTGCTCACTGGTGGAACTTTGGCGGCTGGTGGGTGCGAGGGGTGAGATTCAA 1681
Db 385 TGAGCTTCTGCTCGGAGGAGGCTCGATTGGCGGCGGCGGCAATCCCGTGAATCCAGC 444
Qy 1682 ACCGGTGAATAAGAGTGAAGTGGTGGTGAACCTTTGATGAGACCAATGACTACTCGT---ATG 1738
Db 445 TCGCATCTCCAGAGCAGCTGCTGCAACTTTAAACGAGCGGAGAACGACTACTGTTCCGATG 504
Qy 1739 GGACGAACACCGCTTCCAGGATTGCGACGAGGTGACGGTGTATGTCAATAGCCCGCTGG 1798
Db 505 GGACGAACACCGCTTCTGCTGACTGGGATCGGGTCTATTGTACCAAGACGCGCAATAG 564
Qy 1799 TGTGGGGAGCTGAACCGT 1816
Db 565 TGTGGGAGACTGCTCCAT 582

RESULT 8

US-09-136-574A-46
; Sequence 46, Application US/09136574A
; Patent No. 6294366
; GENERAL INFORMATION:
; APPLICANT: Farrington, Graham K.
; ANDERSON, PAIGE
; GIBBS, MORELAND
; BERGQUIST, PETER
; DANIELS, ROY
; MORGAN, HUGH W.
; WILLIAMS, DIANE P.
; TITLE OF INVENTION: Compositions and Methods for Treating Cellulose Containing Fabrics Using Truncated Cellulase Enzyme Compositions
; NUMBER OF SEQUENCES: 49
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Howson and Howson
; STREET: Spring House Corporate Center, P.O. Box 457
; CITY: Spring House
; STATE: PA
; COUNTRY: USA
; ZIP: 19477
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/136.574A

TELECOMMUNICATION INFORMATION:
TELEPHONE: (716) 263-1304
TELEFAX: (716) 263-1600
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 2712 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
US-09-025-691-4

Query Match 2.9%; Score 65.4; DB 3; Length 2712;
Best Local Similarity 46.9%; Pred. No. 2e-05;
Matches 204; Conservative 0; Mismatches 231; Indels 0; Gaps 0;

Qy 1845 CCCAGCCCAACCCCGTCCCGAGCCGAGCCGACCCAGCCCAAGCCCGAGCTCTCCCCATC 1904
Db 793 CCCCAGACCCAGCCGCGCCGAGCCGCTCCAGAGCCCTTCCGCGCGCGCC 852
Qy 1905 CCGGTCCCGAGCCGAGCCGAGCCCTACGCGCTCCCGCTCCCGGGATGCTCGCGCC 1964
Db 853 GCTGCGCGGCTCGTCCGCGCTCGCGCTCCCGCTCCCGGGATGCTCGCGCC 912
Qy 1965 GCCGAGTGTGCTGCTCGGTGCTGGGTGCTGGGCGCGGCGACGTATGTGTTCAATGATTTG 2024
Db 913 CACCCAGGCGCGGAGCGCGCGGCGGACCCAGCGCTCTTACACCCAGGACGCGACTG 972
Qy 2025 GGTCTCTGGTTTACGGCGAGCGGTACCGTACGAGTACCGGAGCCGCGGCGACGAGCGG 2084
Db 973 GGGCAGCGCTTTCGAGGCGAAGTGGACGGTGAAGAACCCGCGACCGCCCTCCAGCGG 1032
Qy 2085 GTGACGGTGGCGGTGCTGTTTGGTGGGAATCAGACGGTACGAACTACTTGAACACTGC 2144
Db 1033 CTGACCTCGAGTGGGACTTCCCGCGCGGACCAAGGTGACCTCGGCTGGAGCCCGA 1092
Qy 2145 GTTGACCAATCAGTGTGCTGCTGACGCGGACGAACTGAGCTACACAACTGATCCA 2204
Db 1093 CGTACCAACAAACGCGACCACTGGACCGCCCAAGAACAGAGCTGGCGGGGAGCCTCGC 1152
Qy 2205 ACCGGTCACTCGACCACTTCGGATTCAACGGAAGTTACTCAGGAACAAACACCGCAC 2264
Db 1153 CCGCGCGCGCTCGGTCACTTCCGCTTCAACGGCACCGGCGCCCGCTCGGGCTG 1212
Qy 2265 TACACTCACTCCAC 2279
Db 213 CAAGCTCAACGCGC 1227

RESULT 13
US-09-329-234A-6
Sequence 6, Application US/09329234A
Patent No. 6331416
GENERAL INFORMATION:
APPLICANT: Shenai, Ziv
APPLICANT: Shoseyov, Oded
TITLE OF INVENTION: PROCESS OF EXPRESSING AND ISOLATING RECOMBINANT PROTEINS AND RECO
FILE OF INVENTION: PROTEIN PRODUCTS FROM PLANTS, PLANT DERIVED TISSUES OR CULTURED
FILE REFERENCE: 00/20274
CURRENT APPLICATION NUMBER: US/09/329.234A
CURRENT FILING DATE: 2001-09-04
NUMBER OF SEQ ID NOS: 8
SOFTWARE: PatentIn version 3.1
SEQ ID NO 6
LENGTH: 1305
TYPE: DNA
ORGANISM: Artificial sequence
FEATURE:
OTHER INFORMATION: Construct containing Protein L, fused to CBDcex sequence
US-09-329-234A-6

Query Match 2.8%; Score 63; DB 4; Length 1305;
Best Local Similarity 53.4%; Pred. No. 5.4e-05;

Matches 132; Conservative 0; Mismatches 115; Indels 0; Gaps 0;
Qy 2021 ATTGGGGTCTGGGTTTACGGCGACGGTACGGTACGAGTACCGGAGCCCGGGCGACGA 2080
Db 991 AGTGAACACACGGCTTACCGCGAAGCTCACCGTGAAGAACAACGCTCTCCGGTCCGGTAG 1050
Qy 2081 GCGGGTGGACGGTGGCGTGTCTTTGGTGGGAATCAGACGGTACGGAACCTACTGGAACA 2140
Db 1051 ACGGCTGGAGCTCAGGTTCCCGTCCCGCCAGGAGTCAACCGGCTGGAGCT 1110
Qy 2141 CTGCGTTGACCAATCAGTGTGTCATCGTACGGCGACGACCTGAGCTACAAACAGTCA 2200
Db 1111 CGACGGTACGCGAGTCCGGCTCGCGGTGACGGTCCGCAACGCCCTGCAACGGCTCGA 1170
Qy 2201 TCCACCGGGTCACTGCGACCACTTCCGATTCAACGGAAGTTACTCAGGAGAAACACCG 2260
Db 1171 TCCCGCGGGCGGCGACCGCGGAGTTTCGGCTTCAACGGCTTCGACAGGGGACCAACGCGC 1230
Qy 2261 CACCTAC 2267
Db 1231 CGCCGAC 1237

RESULT 14
US-08-118-200-1/c
Sequence 1, Application US/08118200
Patent No. 6197500
GENERAL INFORMATION:
APPLICANT: SUTHERLAND, Grant R
APPLICANT: RICHARDS, Robert J
APPLICANT: SCHLESSINGER, David
APPLICANT: NAGARAJA, Ramalah
APPLICANT: KREMER, Eric J
APPLICANT: YU, Sui
APPLICANT: BAKER, Elizabeth
APPLICANT: MULLEY, John C
APPLICANT: MANDEL, Jean-Louis
APPLICANT: PRITCHARD, Melanie April
APPLICANT: LYNCH, Michael
TITLE OF INVENTION: DNA SEQUENCES RELATED TO ISOLATED
TITLE OF INVENTION: FRAGILE X SYNDROME
NUMBER OF SEQUENCES: 13
CORRESPONDENCE ADDRESS:
ADDRESSEE: BURNS, DOANE, SNECKER & MATHIS
STREET: P.O. Box 1404
CITY: Alexandria
STATE: Virginia
COUNTRY: United States
ZIP: 22313-1404
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/118.200
FILING DATE: 09-SEP-1993
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/802,650
FILING DATE: 05-DEC-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/672,232
FILING DATE: 20-MAR-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/638,518
FILING DATE: 04-JAN-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/966,517
FILING DATE: 23-DEC-1992
ATTORNEY/AGENT INFORMATION:
NAME: Crane-Feury, Sharon E
REGISTRATION NUMBER: 36,113

```

: REFERENCE/DOCKET NUMBER: 020160-164
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (703) 836-6620
: TELEFAX: (703) 836-2021
: INFORMATION FOR SEQ ID NO: 1:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1028 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: DNA (genomic)
: US-08-118-200-1

Query Match          2.7%   Score 62.4; DB 4; Length 1028;
Best Local Similarity 57.0%; Pred. No. 6.8e-05;
Matches 114; Conservative 0; Mismatches 86; Indels 0; Gaps 0;

Qy 1161 CGCGTTCAACGTCTACTACCAAGTTCAACAGCGACCGTCTCGCAGCGAGCGCTCGC 1220
Db 553 CAGTTCACACACCGCTCTCTTTCAGCCCTGCTAGCGCGGAGCGCGCC 494
Qy 1221 AATCAGTACCGGCACATCGCTTCGCCCGTTCGCGAGCTCGAGTCCATCTCTCGCCGTC 1280
Db 493 CCGAAGAGTGGCTCCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCC 434
Qy 1281 TCCGTCCGCGTCTCCGTCCGCGTCTCCGTCCGCGTCTCCGTCCGCGTCTTCGTCCCGCAG 1340
Db 433 GCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCC 374
Qy 1341 CCCGTCTCCGTCTCGTCCG 1360
Db 373 GCCGCGCGCGCGCGCGCGCG 354

RESULT 15
US-08-458-745-1/c
: Sequence 1, Application US/08458745
: Patent No. 6242576
: GENERAL INFORMATION:
: APPLICANT: SUTHERLAND, Grant R
: APPLICANT: RICHARDS, Robert I
: APPLICANT: SCHLESSINGER, David
: APPLICANT: NAGARAJA, Ramaiah
: APPLICANT: KREMER, Eric J
: APPLICANT: YU, Sui
: APPLICANT: BAKER, Elizabeth
: APPLICANT: MULLEY, John C
: APPLICANT: MANDEL, Jean-Louis
: APPLICANT: PRITCHARD, Melanie April
: APPLICANT: LYNCH, Michael
: TITLE OF INVENTION: DNA SEQUENCES RELATED TO ISOLATED
: TITLE OF INVENTION: FRAGILE X SYNDROME
: NUMBER OF SEQUENCES: 13
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: BURNS, DOANE, SWECKER & MATHIS
: STREET: P.O. Box 1404
: City: Alexandria
: STATE: Virginia
: COUNTRY: United States
: ZIP: 22313-1404
: COMPUTER READABLE FORM:
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08458,745
: FILING DATE: 02-JUN-1995
: CLASSIFICATION: 530
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/118,200
: FILING DATE: 09-SEP-1993
: APPLICATION NUMBER: US 07/802,650
```

```

: FILING DATE: 05-DEC-1991
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 07/672,232
: FILING DATE: 20-MAR-1991
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 07/638,518
: FILING DATE: 04-JAN-1991
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 07/966,517
: FILING DATE: 23-DEC-1992
: ATTORNEY/AGENT INFORMATION:
: NAME: Crane-Feury, Sharon E
: REGISTRATION NUMBER: 36,113
: REFERENCE/DOCKET NUMBER: 020160-164
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (703) 836-6620
: TELEFAX: (703) 836-2021
: INFORMATION FOR SEQ ID NO: 1:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1028 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: DNA (genomic)
: US-08-458-745-1

Query Match          2.7%   Score 62.4; DB 4; Length 1028;
Best Local Similarity 57.0%; Pred. No. 6.8e-05;
Matches 114; Conservative 0; Mismatches 86; Indels 0; Gaps 0;

Qy 1161 CGCGTTCAACGTCTACTACCAAGTTCAACAGCGACCGTCTCGCAGCGAGCGCTCGC 1220
Db 553 CAGTTCACACACCGCTCTCTTTCAGCCCTGCTAGCGCGGAGCGCGCC 494
Qy 1221 AATCAGTACCGGCACATCGCTTCGCCCGTTCGCGAGCTCGAGTCCATCTCTCGCCGTC 1280
Db 493 CCGAAGAGTGGCTCCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCC 434
Qy 1281 TCCGTCCGCGTCTCCGTCCGCGTCTCCGTCCGCGTCTCCGTCCGCGTCTTCGTCCCGCAG 1340
Db 433 GCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCC 374
Qy 1341 CCCGTCTCCGTCTCGTCCG 1360
Db 373 GCCGCGCGCGCGCGCGCGCG 354

Search completed: November 13, 2002, 10:50:38
Job time : 118 secs
```


; PRIOR FILING DATE: 1998-11-23
; PRIOR APPLICATION NUMBER: 07/829,461
; PRIOR FILING DATE: 1992-01-31
; PRIOR APPLICATION NUMBER: 07/739,055
; PRIOR FILING DATE: 1991-08-01
; PRIOR APPLICATION NUMBER: 07/561,968
; PRIOR FILING DATE: 1990-08-02
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 7
; LENGTH: 390
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Sequence
; OTHER INFORMATION: having a 120 repeat of ACG flanked by fixed
; OTHER INFORMATION: fragments having NcoI restriction sites.
US-09-790-399-7

Query Match 3.4%; Score 78.4; DB 10; Length 390;
Best Local Similarity 52.8%; Pred. No. 1.3e-09;
Matches 169; Conservative 0; Mismatches 151; Indels 0; Gaps 0;

Qy 1212 GCCGCTCCCAATCAGTACCGGCACATCCCTCCGCGTCCGCGAGCTCGAGTCCATCCTC 1271
Db 386 GACCATCCCAACCATGGCGTCTCGTCTCGTCTCGTCTCGTCTCGTCTCGTCTCGTCT 327
Qy 1272 GTCGCCGCTCCGTCGCGCTCTCGTCTCGGCTCTCCGTCGCGCTCTCGTCTCGGCTCTTC 1331
Db 326 GTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTC 267
Qy 1332 GTCGCCGAGCCCGTCTCGTCTCGTCTCGTCTCGGCTCTCGGTCGAGTGCATACAA 1391
Db 266 GTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTC 207
Qy 1392 CAACAATGATTCGCGCGCGGTGATACCAACACATCAACCGGCTCTCCAGTTGGTGAATAC 1451
Db 206 GTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTC 147
Qy 1452 GGGTCTGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 1511
Db 146 GTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 87
Qy 1512 TGGGTCGTCGACACTGGTGT 1531
Db 86 GTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 67

RESULT 5
US-09-974-300-684
; Sequence 684, Application US/09974300
; Patent No. US20020146721A1
; GENERAL INFORMATION:
; APPLICANT: Berk, Randy M.
; APPLICANT: Clausen, Ib Groth
; TITLE OF INVENTION: Methods For Monitoring Multiple Gene
; TITLE OF INVENTION: Expression
; FILE REFERENCE: 10085.500-US
; CURRENT APPLICATION NUMBER: US/09/974,300
; PRIOR FILING DATE: 2001-10-05
; PRIOR APPLICATION NUMBER: 09/680,598
; PRIOR FILING DATE: 2000-10-06
; PRIOR APPLICATION NUMBER: 60/279,526
; PRIOR FILING DATE: 2001-03-27
; NUMBER OF SEQ ID NOS: 8481
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 684
; LENGTH: 1314
; TYPE: DNA
; ORGANISM: Bacillus licheniformis
US-09-974-300-684

Best Local Similarity 52.9%; Pred. No. 3.4e-07;
Matches 148; Conservative 0; Mismatches 132; Indels 0; Gaps 0;

Qy 1535 ACTGTGACTGGCGCGGATGGGTGTGGGAATATATCCGGCCCTCGTTCGGCTCGGTGAACC 1594
Db 1032 ACTGGACTATGCCCAATCGGCTCGAGCAAAATCAGGCACAAATTCGTTCAATTAATAA 1091
Qy 1595 CGGACACCGCGGCGGACACCTACCTGCGAGTTCGTTCTGCTGCTGCAAGCTTGGCGG 1654
Db 1092 AAGCGTTAAACGAGCAGACAGCTATCTTGAAGTAGGATTTAAAAATGGTACATTGGCG 1151
Qy 1655 CTGGTGGGTTCGACGGTTCAGATTCAAAACCGGGTCAATAAGAGTGAAGTTCGTCGAAC 1714
Db 1152 CGGGGCTCATACTGCGCAATCCAGATCCGCTTTCACAATCAGCGCTGGAGCAATTATG 1211
Qy 1715 ATGAGACCAATGACTACTGCTGATGGGAGCAACACCGCTTCCAGGATTTGGAGCAAGTGA 1774
Db 1212 CCCAAGCGGCGCTATTTCAATTTAAATTCAAACACGCTTTAAAAATACGAAAAAATCA 1271
Qy 1775 CGGTGATGTAATGCGCGCTGCTGGGGGACTGAACC 1814
Db 1272 GCTGTATGAGAACGGAAGCTGATTTGGGCACTGAACC 1311

RESULT 6
US-09-748-033-4
; Sequence 4, Application US/09748033
; Patent No. US2002069431A1
; GENERAL INFORMATION:
; APPLICANT: Broadway, Roxanne M.
; APPLICANT: Gongora, Carmenza E.
; TITLE OF INVENTION: EFFECT OF ENDOCHITINASE AND CHITOBIOSIDASE AND THEIR
; TITLE OF INVENTION: ENCODING GENES ON PLANT GROWTH AND DEVELOPMENT
; FILE REFERENCE: 19603/3091
; CURRENT APPLICATION NUMBER: US/09/748,033
; CURRENT FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/172,003
; PRIOR FILING DATE: 1999-12-23
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4
; LENGTH: 2712
; TYPE: DNA
; ORGANISM: Streptomyces albidoflavus
US-09-748-033-4

Query Match 2.9%; Score 65.4; DB 10; Length 2712;
Best Local Similarity 46.9%; Pred. No. 2.6e-06;
Matches 204; Conservative 0; Mismatches 231; Indels 0; Gaps 0;

Qy 1845 CCCAGCCCAACCCCGTCCCGAGCCCGAGCCCGAGCCCTACCGCTCCCGGAGCCCGCTCC 1904
Db 793 CCCAGACCCGAGCCCGGCGGCGGCTCCAGACCCCTCAGCGCTTCGCGCGGCGCGC 852
Qy 1905 CCCGTCCCGAGCCCGAGCCCGAGCCCTACCGCTCCCGGAGCCCGCTCCCGCTCC 1964
Db 853 GCTGGCGCGCTCGTCGCGCTCGCGGCTCCCTTCGCGGGGATGTCGCGCTCGCGCGCC 912
Qy 1965 GCGAGTGTGTCGTCGCGGTCGCGGTCGCGGCGAGCTATGTGTGAATGATGTTG 2024
Db 913 CACCCAGCCCGGCGGCGGCGGCGGCGGCGGCGGCTTCCTACACGAGCAGCAGGACTG 972
Qy 2025 GGGTCTCGGTTTACGGCGAGGTCAGGTCAGCAATACCGGAGCGGCGGAGCGGCGG 2084
Db 973 GGGCAGCGGCTTCAGGGCAAGTGGAGGCTGAAGAACCACCGGCGGCGGCGGCGGCGG 1032
Qy 2085 GTGACGCTGGCTGGTGGTCTGTTGTTGGGAATCAGACGCTACGAACACTACTGAACAC 2144
Db 1033 CTGGACCTGGAGTGGGACTTCCCGCGGCGGAGCAAGGTCGCTCGGCTGGGAGCGGA 1092
Qy 2145 GTTGACCAATCAGTGTGATGCTGGTGACGGGAGCAACCTGAGCTACCAACAGTGATCCA 2204
Db 1093 CGTACCACCAACCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1152


```
RESULT 10
US-09-974-300-3603
; Sequence 3603, Application US/09974300
; Patent No. US20020146721A1
; GENERAL INFORMATION:
; APPLICANT: Berka, Randy M.
; APPLICANT: Clausen, Ib Groth
; TITLE OF INVENTION: Methods for Monitoring Multiple Gene Expression
; FILE REFERENCE: 10085-500-US
; CURRENT APPLICATION NUMBER: US/09/974,300
; PRIOR FILING DATE: 2001-10-05
; PRIOR APPLICATION NUMBER: 09/680,598
; PRIOR FILING DATE: 2000-10-06
; PRIOR APPLICATION NUMBER: 60/279,526
; PRIOR FILING DATE: 2001-03-27
; NUMBER OF SEQ ID NOS: 8481
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3603
; LENGTH: 300
; TYPE: DNA
; ORGANISM: Bacillus licheniformis
US-09-974-300-3603

Query Match      2.5%; Score 58.2; DB 10; Length 300;
Best Local Similarity 55.1%; Pred. No. 6.6e-05;
Matches 158; Conservative 0; Mismatches 123; Indels 6; Gaps 2;

Qy 935 ACTGGGCCAGACCGGACCTGGGGAACGCAATGGATCAAGGATCATATTCCGGAATGCCG 994
Db 3 ATTGGGGTTATACGAGGCGAGTGGGGGATTCAGTGCATCAGACAGACAGATTTGCGATGGGA 62

Qy 995 CAGCGATCGGCAAGCCGACCAATTCGGAAGAAATTCGGGTGGCAGACACCGGACCGC---G 1051
Db 63 AAGNATCGGCAAGCCGTCGTTTTAGAGAGATCGGGTATCAGTATAGTATCAGGAAGGG 122

Qy 1052 ATTCCGCTATCAGACGTGGACCCAGACTGTGCGGTACGAAGCGGTGAACAGCGCTTGAAC 1111
Db 123 ACTACGCTACAGAACCTGCTGTTAACTCATAGAAAGCAGAGCGGTGCGGGCAGCCAAAT 182

Qy 1112 TCTGATGCTCGCTGGGA---ATCTCAAGCCGACCCATATCCGAACATATACGCGCTTCA 1168
Db 183 TTTGGATTTTACCGGATTCAGGATCAGGACGACCGGACCTTTATCCGGACTATACCGGTTTC 242

Qy 1169 ACGTCTACTACCCAAAGTTCAACAGCAGCGCTCCCTCGGCGAGCGGCG 1215
Db 243 GGATCGTTATCCGAGCTCTCCCGCTTCGTCATTTTCAGACGACGCG 289

RESULT 11
US-09-860-846-34/c
; Sequence 34, Application US/09860846
; Patent No. US20020164742A1
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
; FILE REFERENCE: 600.438US1
; CURRENT APPLICATION NUMBER: US/09/860,846
; CURRENT FILING DATE: 2001-05-18
; PRIOR APPLICATION NUMBER: 09/105,537
; PRIOR FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 34
; LENGTH: 4689
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
US-09-860-846-34

Query Match      2.5%; Score 57.4; DB 10; Length 4689;
Best Local Similarity 46.4%; Pred. No. 0.00023;
Matches 187; Conservative 0; Mismatches 216; Indels 0; Gaps 0;

Qy 1209 CGAGGCGCTCGCAATCACTACCGGCACATCGCCTCCGCGGTCCGCGAGCTCGAGTCCATC 1268
Db 3912 CGCGTCGTGGCGCCCGCCCATCGCCTCGCGGCCACAGAGGCCACCGGACCGAGGT 3853

Qy 1269 CTCGTCGCGCGTCTCGTCCGCTCTCCGTCGCGGTCTCCGTCGCGGTCTCCGTCGCGGT 1328
Db 3852 CGCGTCTCGCGCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3793

Qy 1329 TTCGTGCGCGAGCCGCTCTCGTCTCGTTCGCGCGGTTCGCGGTGCGGTGCAAGTTCAGTA 1388
Db 3792 CGCGTAGAGCCCTTGGCTGCGCGCTGCCCGACAGACCCCGCGGTTCGAGGAGTAGAGACGAA 3733

Qy 1389 CAAGAACAATGATTTCGCGCGCGGTGATACACAGATCAACACGGGTCTCCGATTTGGTCAA 1448
Db 3732 GCGCTCCAGCGAGTGCCTCGCGAGCAGAGTTCGTTCGAGGACCTCGCGCGCGCTCTGCTTCG 3673

Qy 1449 TACGGGTCTGTCGTCGCTGGATTTCGTCACGCTGACGCTGCGGTACTGTTTCACCCGGA 1508
Db 3672 GCCAGGATCGGGCGATGTCCTCGCGCGCGGTGACGTCCAGCGGATCGCGCCCGGTCG 3613

Qy 1509 TCGTGGTCTCGACACTGGTCTACAACGTGCTACTGGCGCGGATGGGTGTGGGAATAT 1568
Db 3612 GCCGCGGTGTGGACGACGCGGTGACGGGCTCTCGCGGGGATGCGCTCGAGCAGGCT 3553

Qy 1569 CCGGCGCTCTGTCGCTCGGTGAACCCGCGGCGACGCGCGGCGG 1611
Db 3552 GCGATGCGGTGGGGTTCGCGGACGCTCGCAGCGCGGATGGTG 3510

RESULT 12
US-09-861-289-34/c
; Sequence 34, Application US/09861289
; Patent No. US20020110897A1
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
; FILE REFERENCE: 600.438US1
; CURRENT APPLICATION NUMBER: US/09/861,289
; CURRENT FILING DATE: 2001-05-18
; PRIOR APPLICATION NUMBER: 09/105,537
; PRIOR FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 34
; LENGTH: 4689
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
US-09-861-289-34

Query Match      2.5%; Score 57.4; DB 10; Length 4689;
Best Local Similarity 46.4%; Pred. No. 0.00023;
Matches 187; Conservative 0; Mismatches 216; Indels 0; Gaps 0;

Qy 1209 CGAGGCGCTCGCAATCACTACCGGCACATCGCCTCCGCGGTCCGCGAGCTCGAGTCCATC 1268
Db 3912 CGCGTCGTGGCGCCCGCCCATCGCCTCGCGGCCACAGAGGCCACCGGACCGAGGT 3853

Qy 1269 CTCGTCGCGCGTCTCGTCCGCTCTCCGTCGCGGTCTCCGTCGCGGTCTCCGTCGCGGT 1328
Db 3852 CGCGTCTCGCGCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3793

Qy 1329 TTCGTGCGCGAGCCGCTCTCGTCTCGTTCGCGCGGTTCGCGGTGCGGTGCAAGTTCAGTA 1388
Db 3792 CGCGTAGAGCCCTTGGCTGCGCGCTGCCCGACAGACCCCGCGGTTCGAGGAGTAGAGACGAA 3733

Qy 1389 CAAGAACAATGATTTCGCGCGCGGTGATACACAGATCAACACGGGTCTCCGATTTGGTCAA 1448
```

Db 3732 GCGCTCCAGCGAGTGCCTGGGAGGAGGTCGTCGAGGACCTCGGCGCGCTCGTCTTCGC 3673
Qy 1449 TACGGGTCTGCTCGGTGGGATTTGTCACGCTGACGCTGCGGTACTGTTCCACCGGGA 1508
Db 3672 GCCCAGGATGGGGGATGTCCTCGGGCCGCTGACGTCCAGCGGATCGCCCGCGGTGC 3613
Qy 1509 TGGTGGGTCTGACACTGTTGTAACAATGTACTGGCGCGGATGGGGTGTGGGAATAT 1568
Db 3612 GCGCGGTGTGGACGAGCGCGGTGAGGGCGTCTCGCGGGGATGGCTCAGAGAGGT 3553
Qy 1569 CCGCGCTCTGCTCGGTGCTCAACCGCGCGGACCGGAGCGG 1611
Db 3552 GCGCATGCGGTGGGGTTCGCGGACGTCGCGCGCGGATGGTG 3510

RESULT 13
US-09-860-846-5/c
: Sequence 5, Application US/09860846
: Patent No. US20020164742A1
: GENERAL INFORMATION:
: APPLICANT: Sherman, D.H.
: APPLICANT: Liu, H.
: APPLICANT: Xue, Y.
: APPLICANT: Zhao, L.
: TITLE OF INVENTION: DNA encoding methymycin and pikromycin
: FILE REFERENCE: 600.438US1
: CURRENT APPLICATION NUMBER: US/09/860.846
: PRIOR FILING DATE: 2001-05-18
: PRIOR APPLICATION NUMBER: 09/105,537
: PRIOR FILING DATE: 1998-06-26
: NUMBER OF SEQ ID NOS: 43
: SOFTWARE: FastSeq for Windows Version 3.0
: SEQ ID NO 5
: LENGTH: 36778
: TYPE: DNA
: ORGANISM: Streptomyces venezuelae
US-09-860-846-5

Query Match 2.5%; Score 57.4; DB 9; Length 36778;
Best Local Similarity 46.4%; Pred. No. 0.00044;
Matches 187; Conservative 0; Mismatches 216; Indels 0; Gaps 0;

Qy 1209 CGAGGCGCTCCCAATCAGTACCGGCACATCGCCTCCGCGCTCGCGGAGCTCGAGTCCATC 1268
Db 30902 CCGCTGCTGGCGCGCGCCGATGCGCTCGCGCGCCAGAGCGCCAGGCGACGAGGT 30843
Qy 1269 CTCGTGCGCGCTCTCGCTCGCGCTCTCCGTCGCGCTCTCCGTCGCGGTCTCCGTCGCGGT 1328
Db 30842 CCGCTCTCG 30783
Qy 1329 TTCGTGCGCGGAGCGCGCTCTCGCTCGCTCGCGCGGTGTGGGTGGGTGAAGTGCAGTA 1388
Db 30782 CCGCTAGACGCGCTGCTGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 30723
Qy 1389 CAAGAACAAATGATTCGCGCGCGCGGTGATAACACAGATCAAAACCGGCTCTCCAGTTGTTGAA 1448
Db 30722 GCGCTCAGCGAGTGCCTCGCGGAGAGGTGCTCGAGGACCTCGCGCGCGCTCTTCGC 30663
Qy 1449 TACGGGTCTGCTCGGTGGGATTTGTCAGCGGTGACGCTGCGGTGCGGTACTGTTTCCACCGGGA 1508
Db 30662 GCCCAGGATGGGGGATGCTCTCGCGCGCGGTGACGTCCAGCGGATCGCCCGCGGTGC 30603
Qy 1509 TGGTGGGTCTGACACTGGTGTACAACGTGACGTCGCGGTGGGTGGGTGAAGTGAATAT 1568
Db 30602 CCGCGCGGTGTGGACGAGCGGTGAGGGGCTCTCGCGCGGATGGCTCAGAGGAGGT 30543
Qy 1569 CCGCGCTCTGCTCGGTGCTGTAACCGCGGACCGCGGAGCGCGG 1611
Db 30542 CCGCATGCGGTGGGGTTCGCGGACGTCGCGCGCGGATGGTG 30500

: Sequence 5, Application US/09861289
: Patent No. US20020110897A1
: GENERAL INFORMATION:
: APPLICANT: Sherman, D.H.
: APPLICANT: Liu, H.
: APPLICANT: Xue, Y.
: APPLICANT: Zhao, L.
: TITLE OF INVENTION: DNA encoding methymycin and pikromycin
: FILE REFERENCE: 600.438US1
: CURRENT APPLICATION NUMBER: US/09/861.289
: PRIOR FILING DATE: 2001-05-18
: PRIOR APPLICATION NUMBER: 09/105,537
: PRIOR FILING DATE: 1998-06-26
: NUMBER OF SEQ ID NOS: 43
: SOFTWARE: FastSeq for Windows Version 3.0
: SEQ ID NO 5
: LENGTH: 36778
: TYPE: DNA
: ORGANISM: Streptomyces venezuelae
US-09-861-289-5

Query Match 2.5%; Score 57.4; DB 10; Length 36778;
Best Local Similarity 46.4%; Pred. No. 0.00044;
Matches 187; Conservative 0; Mismatches 216; Indels 0; Gaps 0;

Qy 1209 CGAGGCGCTCCCAATCAGTACCGGCACATCGCCTCCGCGCTCGCGGAGCTCGAGTCCATC 1268
Db 30902 CCGCTGCTGGCGCGCGCCGATGCGCTCGCGCGCCAGAGCGCCAGGCGACGAGGT 30843
Qy 1269 CTCGTGCGCGCTCTCGCTCGCGCTCTCCGTCGCGCTCTCCGTCGCGGTCTCCGTCGCGGT 1328
Db 30842 CCGCTCTCG 30783
Qy 1329 TTCGTGCGCGGAGCGCGCTCTCGCTCGCTCGCGCGGTGTGGGTGGGTGAAGTGCAGTA 1388
Db 30782 CCGCTAGACGCGCTGCTGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 30723
Qy 1389 CAAGAACAAATGATTCGCGCGCGCGGTGATAACACAGATCAAAACCGGCTCTCCAGTTGTTGAA 1448
Db 30722 GCGCTCAGCGAGTGCCTCGCGGAGAGGTGCTCGAGGACCTCGCGCGCGCTCTTCGC 30663
Qy 1449 TACGGGTCTGCTCGGTGGGATTTGTCAGCGGTGACGCTGCGGTGCGGTACTGTTTCCACCGGGA 1508
Db 30662 GCCCAGGATGGGGGATGCTCTCGCGCGCGGTGACGTCCAGCGGATCGCCCGCGGTGC 30603
Qy 1509 TGGTGGGTCTGACACTGGTGTACAACGTGACGTCGCGGTGGGTGGGTGAAGTGAATAT 1568
Db 30602 CCGCGCGGTGTGGACGAGCGGTGAGGGGCTCTCGCGCGGATGGCTCAGAGGAGGT 30543
Qy 1569 CCGCGCTCTGCTCGGTGCTGTAACCGCGGACCGCGGAGCGCGG 1611
Db 30542 CCGCATGCGGTGGGGTTCGCGGACGTCGCGCGCGGATGGTG 30500

RESULT 15
US-09-772-105-48/c
: Sequence 48, Application US/09772105
: Patent No. US20010029015A1
: GENERAL INFORMATION:
: APPLICANT: Ozelius, Laurie J.
: APPLICANT: Breakfield, Xandra O.
: TITLE OF INVENTION: TORSIN, TORSIN-RELATED GENES, AND
: TITLE OF INVENTION: METHODS OF DETECTING NEURONAL DISEASES
: FILE REFERENCE: 0838.1001009
: CURRENT APPLICATION NUMBER: US/09/772.105
: CURRENT FILING DATE: 2001-01-26
: PRIOR APPLICATION NUMBER: US 09/218,363
: PRIOR FILING DATE: 1998-12-22
: PRIOR APPLICATION NUMBER: US 09/099,454
: PRIOR FILING DATE: 1998-06-18
: PRIOR APPLICATION NUMBER: US 60/050,244
: PRIOR FILING DATE: 1997-06-19
: NUMBER OF SEQ ID NOS: 90

```
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 48
; LENGTH: 283
; TYPE: DNA
; ORGANISM: Unknown
; FEATURE:
; OTHER INFORMATION: cDNA clone of DYT1 Intron 1
; NAME/KEY: misc_feature
; LOCATION: (1)..(283)
; OTHER INFORMATION: n = A,T,C or G
US-09-772-105-48

Query Match      2.4%; Score 55; DB 10; Length 283;
Best Local Similarity 63.0%; Pred. No. 0.00036;
Matches 85; Conservative 0; Mismatches 50; Indels 0; Gaps 0;

Qy 1827 CCCAGCCCCACACCCAGCCCGCCGAGCCCAACCCCGTCCCGGAGCCGCGACCCCAAG 1886
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 238 CCCACCCCGCAGCCCGCCGCGCGTCCATCCCGGAGCCCGCCAGCCCAACCCGCC 179
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 1887 CCCAGGTCTCTCCCGCATCCCGGTCCCGGAGCCCGCCAGCCCGCCAGCCCTACGGCGTCCCGGTC 1946
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 178 CTAGTGCCATCGCCCGCAGCCCGCCAGCCCGCCAGCCCGCCGCTCCAGCCCTAGTCTAGCCCGGC 119
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 1947 GCCGAGCCCGTCCGCC 1961
    ||| ||| |||
Db 118 CCTGGTGCATCCCC 104
```

Search completed: November 13, 2002, 11:49:40
Job time : 179 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 13, 2002, 10:49:26 ; Search time 33.6746 Seconds
(without alignments)
1483.879 Million cell updates/sec

Title: US-09-917-378-3

Perfect score: 2079

Sequence: 1 APAGFVTASGQFVNLGLPY.....YPSSTATVLASEALAISTG 375

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 20000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A_Geneseq_101002.*

```
1: /SID52/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2: /SID52/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3: /SID52/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4: /SID52/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5: /SID52/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6: /SID52/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7: /SID52/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8: /SID52/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9: /SID52/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10: /SID52/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11: /SID52/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12: /SID52/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13: /SID52/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14: /SID52/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15: /SID52/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16: /SID52/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17: /SID52/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18: /SID52/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19: /SID52/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20: /SID52/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21: /SID52/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22: /SID52/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
23: /SID52/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	829.5	39.9	666	19 AAW34992	Thermotoga neapoli
2	812	39.1	663	19 AAW49872	Thermotoga maritim
3	812	39.1	680	19 AAW35005	Thermotoga maritim
4	806	38.8	680	18 AAW34564	Thermotoga maritim
5	805	38.7	680	19 AAW49868	Thermotoga maritim
6	543.5	26.1	431	21 AAG27501	Arabidopsis thalia
7	543.5	26.1	431	21 AAG42172	Arabidopsis thalia
8	543.5	26.1	431	23 ABB93972	Herbicidally activ
9	543.5	26.1	442	21 AAG42171	Arabidopsis thalia
10	543.5	26.1	443	21 AAG27500	Arabidopsis thalia

11	537	25.8	408	23 ABB92275	Herbicidally activ
12	532.5	25.6	416	22 AAG79297	Amino acid sequenc
13	517.5	24.9	431	23 ABB90809	Herbicidally activ
14	516	24.8	414	23 ABB92274	Herbicidally activ
15	498	24.0	448	23 ABB93325	Herbicidally activ
16	495	23.8	431	23 ABB93133	Herbicidally activ
17	493	23.7	427	21 AAY93441	Amino acid sequenc
18	477	22.9	375	21 AAG27502	Arabidopsis thalia
19	477	22.9	375	21 AAG42173	Arabidopsis thalia
20	468	22.5	351	23 ABB92490	Herbicidally activ
21	459	22.1	415	22 AAE00417	Lycoposicon escul
22	426.5	20.5	369	20 AAY14150	Protein encoded by
23	411.5	19.8	403	23 ABB91820	Herbicidally activ
24	386	18.6	312	21 AAG07279	Arabidopsis thalia
25	347.5	16.7	245	21 AAG07280	Arabidopsis thalia
26	346.5	16.7	230	21 AAG07281	Arabidopsis thalia
27	323	15.5	264	21 AAG47479	Arabidopsis thalia
28	323	15.5	318	21 AAG47478	Arabidopsis thalia
29	315	15.2	264	21 AAG22457	Arabidopsis thalia
30	315	15.2	318	21 AAG22456	Arabidopsis thalia
31	261	12.6	236	21 AAG30018	Arabidopsis thalia
32	261	12.6	236	21 AAG22458	Arabidopsis thalia
33	261	12.6	236	21 AAG47480	Arabidopsis thalia
34	238	11.4	231	21 AAG20019	Arabidopsis thalia
35	201	9.7	210	21 AAG20020	Arabidopsis thalia
36	181.5	8.7	521	22 AAB48786	Acidothermus cellu
37	181.5	8.7	562	21 AAY69508	Acidothermus cellu
38	179.5	8.6	358	19 AAW39262	A. cellulolyticus
39	179.5	8.6	521	17 AAR89927	A. cellulolyticus
40	175.5	8.6	562	23 AAB79549	A. cellulolyticus
41	175.5	8.4	521	22 AAB48787	Acidothermus cellu
42	173.5	8.3	521	22 AAB48788	Acidothermus cellu
43	134	6.4	1684	12 AAR14948	Bacterial amylase
44	127	6.1	425	22 ABB09060	Thermus caldophilu
45	124.5	6.0	711	19 AAW55103	Streptococcus pneu

ALIGNMENTS

RESULT 1

AAW34992

ID AAW34992 standard; Protein: 666 AA.

AC AAW34992;

XX

XX

DT 21-MAY-1998 (first entry)

XX

DE Thermotoga neapolitana endoglucanase.

XX

KW Endoglucanase; cellulase; carboxymethylcellulose; cellulose;

KW biomass; beta-1,4-glycosidic bond; hydrolysis; saccharification;

KW thermostable enzyme; thermophilic; glycosidase.

XX

OS Thermotoga neapolitana (Clone 56GP1).

XX

PN WO9744361-A1.

PD 27-NOV-1997.

XX

XX 22-MAY-1997; 97WO-US08793.

PF

PR 22-MAY-1996; 96US-0651572.

XX

PA (RECO-) RECOMBINANT BIOCATALYSIS INC.

XX

PI Lam DE, Mathur EJ;

XX

DR WPI: 1998-018435/02.

DR N-PSDB; AAT94200.

XX

PT Endoglucanase(s), preferably form archaeal bacterium, AEP11 la -

useful to degrade carboxymethylcellulose and hydrolyse of

```
PT  beta-1,4-glycosidic bonds in cellulose
XX
PS  Claim 1: Fig 1H: 164pp; English.
XX
CC  This protein comprises an endoglucanase of Thermotoga neapolitana
CC  (Clone 56GP1) that is capable of degrading carboxymethylcellulose
CC  and of hydrolysing the beta-1,4-glycosidic bonds in cellulose. It
CC  has homology to an endoglucanase of archaeobacterium AEP11a (see
CC  AAW34985). It can be produced from native cells or from recombinant
CC  host cells, especially prokaryotic host cells transformed with a
CC  plasmid or virus-derived vector including the endoglucanase DNA
CC  (see AAT94200). 24 Endoglucanases (see AAW34986-W35008) are claimed.
CC  They can be used to degrade cellulose for the conversion of plant
CC  biomass into fuels and chemicals, for use in detergents, textiles,
CC  animal feed, waste treatment, and in the fruit juice and brewing
CC  industries for the clarification and extraction of juices.
XX
SQ  Sequence 666 AA:
      Query Match      39.9%; Score 829.5; DB 19; Length 666;
      Best Local Similarity 43.5%; Pred. No. 4e-60;
      Matches 173; Conservative 63; Mismatches 119; Indels 43; Gaps 13:
QY  1  APAGFVTASGQFVLNGLPYRYGGTNNYLSYQSHADVDVLAQAQAMNLSVIRTWGFI 60
DB  1  ANSDFVKVNGRFLNGEERFVGSNNYMHYKSNRMIDSVLESAMKGVKVLRIWGF-- 76
QY  61  IGSIDGSGVPTIDGNKNGFYQYWDPTGCA---PAYNDGPTGLOGLDYAIAASAAHGLRVI 117
DB  77  ---LDGESYCRDKN-----TYMHPAPGVFCLPGCTNAQDGERLDYTVAKAKELGKLI 127
QY  118  VVLTDNKEFGGMDQYDKWGLPYHDNFYTDPTQQAQYKNVNHLLNRVNSITGVYKND 177
DB  128  IVLVNNMDDFGGMQYVRFWFGIHHDFYRNEKIKEEYKVVSLINRVNTYTGVPYREE 187
QY  178  PTIFAWELANEPRCVSGSGLTPTSGCTQATIVNVVDQMSAYKSIDPNHMYSVGDGEF-- 235
DB  188  PTIFAWELANEPRC-----ETDKSGN-----TLVVEWVEMSAYIKSLDPNHLVAVGDEGFN 239
QY  236  -YIGSTGSG---WPYNDPDCGVDDNALLRVKNIIDFGTYHLYPNYWGON----ADWGTOW 287
DB  240  NYEGFRPYGGEAWAYNGWS-GVDWKRLLEIETVDFGTFHLYPSHWGVSPENYAOWGAKW 298
QY  288  IKDHIAAAGKPTILEEFG--WQTP-DRDSVYOTVTTQVTRTNGEAGWNFMMLAGNVNG 344
DB  299  IEDHIKIAKEVGKPVLEEVGIPKSAVNRVAIKLWNLNDLVNLLGGCAMPFMLAGIGEG 358
QY  345  QP-----YPNYDGFNVYPSSTATVLAASE-ALAISTG 375
DB  359  WDRDEKGYPDYDGFRIVNDSEAKLIREYAKLFSTG 396
RESULT 2
AAW49872
ID  AAW49872 standard; Protein: 563 AA.
XX
AC  AAW49872:
XX
DT  21-DEC-1998 (first entry)
XX
DE  Thermotoga maritima MSB8-6GP2 glycosidase.
XX
KW  Glycosidase; MSB8-6GP2; thermostable enzyme; oligosaccharide;
KW  glucose; sugar; baking; textile; detergent; beta-galactosidase.
XX
OS  Thermotoga maritima strain MSB8-6GP2.
XX
PN  WO9824799-A1.
XX
PD  11-JUN-1998.
XX
PF  08-DEC-1997; 97WO-US22623.
XX
```

```
PR  10-OCT-1997; 97US-0949026.
XX  06-DEC-1996; 96US-0056916.
XX  (DIVE-) DIVERSA CORP.
XX  Bylina EJ, Lam DE, Mathur EJ, Swanson RV;
XX  WPI: 1998-362407/31.
XX  N-PSDB: AAW36921.
XX
CC  Glycosidase enzymes from organisms of the genera Staphylothermus,
CC  Pyrococcus and Thermococcus - for deriving sugar from
CC  oligosaccharides, useful in the e.g. food processing, textile or
CC  baking industries
XX
XX  Claim 1: Fig 15a-d; 92pp; English.
XX
CC  This is the amino acid sequence of glycosidase MSB8-6GP2, deduced
CC  from a polynucleotide (see AAW36921) of clone 6GP2 of Thermotoga
CC  maritima MSB8 that grows optimally at 85 degC in high salt medium.
CC  The invention provides 18 polynucleotides (see AAW36907-24) coding
CC  for thermostable glycosidases (see AAW49858-75) having glucosidase,
CC  alpha-galactosidase, beta-galactosidase, beta-mannosidase,
CC  beta-mannanase, endoglucanase or pullulanase activity. Vectors and
CC  host cells are also claimed. A method is provided for producing
CC  the enzymes by recombinant techniques. A claimed method for
CC  generating glucose from soluble cell oligosaccharides comprises
CC  contacting a sample (selected from dairy products, fruit juice,
CC  detergent, textile, guar gum, animal feed, plant biomass or waste
CC  product) containing oligosaccharides (selected from maltose,
CC  cellobiose, lactose, sucrose, raffinose, stachyose, verbascose,
CC  cellulose, starch, amylose, glycogen, disaccharides, polysaccharides
CC  and pullulan) with one of the claimed glycosidases such that glucose
CC  is produced.
XX
SQ  Sequence 663 AA:
      Query Match      39.1%; Score 812; DB 19; Length 663;
      Best Local Similarity 42.8%; Pred. No. 1.1e-58;
      Matches 169; Conservative 61; Mismatches 121; Indels 44; Gaps 13:
QY  5  FVTASGGFVLNGLPYRYGGTNNYLSYQSHADVDVLAQAQAMNLSVIRTWGFI 64
DB  17  FVKVENGKFAALNGKEFRFIGSNYYMHYKSNMIDSVLESARDMGIKVLRIWGF----L 71
QY  65  DGSVPTIDGNKNGFYQYWDPTGAPYNDG----PTGLOGLDYAIAASAAHGLRVI 120
DB  72  DGSYCRDKN-----TYMHPPEPGVFGVPEGISNAQSGFERLDYTVAKAKELGKLI 125
QY  121  TNDWKEFGGMDQYDKWGLPYHDNFYTDPTQQAQYKNVNHLLNRVNSITGVYKNDPTI 180
DB  126  VNNMDDFGGMQYVRFWFGTTHDDFYRDEKIKEEYKVVSLFVNLHVNTYTGVPYREEPTI 185
QY  181  FAWELANEPRCVSGSGLTPTSGCTQATIVNVVDQMSAYKSIDPNHMYSVGDGEF---YI 237
DB  186  MAWELANEPRC-----ETDKSGN-----TLVVEWVKEMSSYIKSLDPNHLVAVGDEGFSNYE 237
QY  238  GSTGSG---WPYNDPDCGVDDNALLRVKNIIDFGTYHLYPNYWGON----ADWGTOWIKD 290
DB  238  GFKPYGGEAWAYNGWS-GVDWKRLLEIETVDFGTFHLYPSHWGVSPENYAOWGAKWIED 296
QY  291  HIAAAGKPTILEEFG--WQTP-DRDSVYOTVTTQVTRTNGEAGWNFMMLAGNVNGOP- 346
DB  297  HIKAKEIGKPVLEEVGIPKSAVNRVAIKLWNLNDLVNLLGGCAMPFMLAGIGEGSDR 356
QY  347  -----YPNYDGFNVYPSSTATVLAASE-ALAISTG 375
DB  357  DERGYYPDYDGFRIVNDSDSEAKLIREYAKLFNTG 391
RESULT 3
AAW35005
ID  AAW35005 standard; Protein: 680 AA.
```


Db 143 VNNMDDFGGNNQYVWFEGGTHDDFKIKEYKKYVFLVNHVNTYTCVYREPTI 202
QY 181 FAMELANEPRCVSGSLPTSCCTCTOATIVNMVDMQSAVKSIDPNHMSVSGDEGF--YI 237
Db 203 IAMELANEPRC-----ETDKSGN-----TLVEWVKEMSSYIKSLDPNHLVAVGDEGFFSNYE 254
QY 238 GSTOGSG---WPYNDPSGDVNNALLRVKNIDFGTYHLYPNYWGON----ADMGTOWIKD 290
Db 255 GFKPYGGEAEWAYNG--SSGVDMKKLLSIETVDFCTHLYPSHMGVSPENYAQMGAKKWIED 313
QY 291 HIANAAIGKPTILEEFG--WQTP-DRDSVYQWTQTVRTNNGEAGNFWMLAGNVNQP- 346
Db 314 HIKIAKEIGKPVVLEEYGIKPSAPVNTAIVRLWDLVYDLGGDGMFWMLAGIGEGSDR 373
QY 347 -----YPNYGDFNVYPSSTATVLAASE-ALAISTG 375
Db 374 DERGYYPDYDGFRIVNDSDPEAELIREYAKLFNTG 408

RESULT 5
AAW49868
ID AAW49868 standard; Protein: 680 AA.
AC AAW49868;
XX
DT 21-DEC-1998 (first entry)
DE Thermotoga maritima 6CP2 beta-mannanase.
DE
KW Glycosidase: 6CP2: thermostable enzyme; oligosaccharide; glucose;
KW sugar; baking; textile; detergent; beta-mannanase.
XX
OS Thermotoga maritima strain 6CP2.
XX
PN WO9824799-A1.
XX
PD 11-JUN-1998.
XX
PF 08-DEC-1997: 97WO-US22623.
XX
PR 10-OCT-1997: 97US-0949026.
PR 06-DEC-1996: 96US-0056916.
XX
PA (DIVE-) DIVERSA CORP.
XX
PI Bylina EJ, Lam DE, Mathur EJ, Swanson RV.
XX
DR WPI: 1998-362407/31.
DR N-PSDB; AAV36917.
XX
PT Glycosidase enzymes from organisms of the genera Staphylothermus,
PT Pyrococcus and Thermococcus - for deriving sugar from
PT oligosaccharides, useful in the e.g. food processing, textile or
PT baking industries
XX
PS Claim 1: Fig 11a-d; 92pp; English.

CC This is the amino acid sequence of beta-mannanase 6CP2, deduced
CC from an isolated polynucleotide (see AAV36917) of a Thermotoga maritima
CC clone (6CP2) that grows optimally at 85 degC in high salt medium.
CC The sequence shows 56% amino acid identity to a mannanase of
CC Aspergillus aculeatus. The invention provides 18 polynucleotides
CC (see AAV36907-24) coding for thermostable glycosidases (see AAW49858-75)
CC having glucosidase, alpha-galactosidase, beta-galactosidase,
CC beta-mannosidase, beta-mannanase, endoglucanase or pullulanase
CC activity. Vectors and host cells are also claimed. A method is
CC provided for producing the enzymes by recombinant techniques. A
CC claimed method for generating glucose from soluble cell
CC oligosaccharides comprises contacting a sample (selected from dairy
CC products, fruit juice, detergent, textile, guar gum, animal feed,
CC plant biomass or waste product) containing oligosaccharides
CC (selected from maltose, cellobiose, lactose, sucrose, raffinose,

CC stachyose, verbascose, cellulose, starch, amylose, glycogen,
CC disaccharides, polysaccharides and pullulan) with one of the
CC claimed glycosidases such that glucose is produced.
XX
SQ Sequence 680 AA;
Query Match 38.7%; Score 805; DB 19; Length 680;
Best Local Similarity 42.5%; Pred. No. 4, 4c-58;
Matches 168; Conservative 61; Mismatches 122; Indels 44; Gaps 13;
QY 5 FVTASGGQFVLNGLPYRYGGTNNYYLSYQSHADVDDVLAKAQAAMNLSVIRTWGFIIDIGSL 64
Db 34 FVKVENGKFAALNGKEFRFGLGNNYYMHYKSNMGIDSVLESARDMGIKVLRWGF-----L 88
QY 65 DGSVPTTDGKNKNGFYFOWDPSTCAPAYNDG-----PTGLOGLDYAIAASAAHGLRVVL 120
Db 89 DGSYCRDKN-----TYMHPEPCVFGVPEGISNAQSGFERLDYTVAKAKELGKLVIVL 142
QY 121 TNDMKFEGGMDQYDKWYCLPYHDNFYTDPRTOQAYKNNVHLLNRVANSITGVTKNDPTI 180
Db 143 VNNWDDFGGNNQYVWFEGGTHDDFKIKEYKKYVFLVNHVNTYTCVYREPTI 202
QY 181 FAMELANEPRCVSGSLPTSCCTCTOATIVNMVDMQSAVKSIDPNHMSVSGDEGF--YI 237
Db 203 MAMELANEPPC-----ETDKSGN-----TLVEWVKEMSSYIKSLDPNHLVAVGDEGFFSNYE 254
QY 238 GSTOGSG---WPYNDPSGDVNNALLRVKNIDFGTYHLYPNYWGON----ADMGTOWIKD 290
Db 255 GFKPYGGEAEWAYNGS--GVDWKKLLSIETVDFCTHLYPSHMGVSPENYAQMGAKKWIED 313
QY 291 HIANAAIGKPTILEEFG--WQTP-DRDSVYQWTQTVRTNNGEAGNFWMLAGNVNQP- 346
Db 314 HIKIAKEIGKPVVLEEYGIKPSAPVNTAIVRLWDLVYDLGGDGMFWMLAGIGEGSDR 373
QY 347 -----YPNYGDFNVYPSSTATVLAASE-ALAISTG 375
Db 374 DERGYYPDYDGFRIVNDSDPEAELIREYAKLFNTG 408

RESULT 6
AAG27501
ID AAG27501 standard; Protein: 431 AA.
XX
AC AAG27501;
XX
DT 17-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 32363.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000: 2000EP-0301439.
XX
PR 25-FEB-1999: 99US-0121825.
PR 05-MAR-1999: 99US-0123180.
PR 09-MAR-1999: 99US-0123548.
PR 23-MAR-1999: 99US-0125788.
PR 29-MAR-1999: 99US-0126264.
PR 29-MAR-1999: 99US-0126785.
PR 01-APR-1999: 99US-0127462.
PR 06-APR-1999: 99US-0128234.
PR 08-APR-1999: 99US-0128714.
PR 16-APR-1999: 99US-0129845.
PR 19-APR-1999: 99US-0130077.
PR 21-APR-1999: 99US-0130449.

PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142134.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 23-APR-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.

PR	14-OCT-1999;	99US-0159637.	
PR	14-OCT-1999;	99US-0159638.	
PR	18-OCT-1999;	99US-0159584.	
PR	21-OCT-1999;	99US-0160741.	
PR	21-OCT-1999;	99US-0160767.	
PR	21-OCT-1999;	99US-0160768.	
PR	21-OCT-1999;	99US-0160770.	
PR	21-OCT-1999;	99US-0160814.	
PR	21-OCT-1999;	99US-0160815.	
PR	22-OCT-1999;	99US-0160980.	
PR	22-OCT-1999;	99US-0160981.	
PR	22-OCT-1999;	99US-0160988.	
PR	25-OCT-1999;	99US-0161404.	
PR	25-OCT-1999;	99US-0161405.	
PR	25-OCT-1999;	99US-0161406.	
PR	26-OCT-1999;	99US-0161359.	
PR	26-OCT-1999;	99US-0161360.	
PR	26-OCT-1999;	99US-0161361.	
PR	28-OCT-1999;	99US-0161920.	
PR	28-OCT-1999;	99US-0161992.	
PR	28-OCT-1999;	99US-0161993.	
PR	29-OCT-1999;	99US-0162142.	
Query Match 26.1%; Score 543.5; DB 21; Length 431;			
Best Local Similarity 35.5%; Pred. No. 1.le-36;			
Matches 139; Conservative 49; Mismatches 139; Indels 65; Gaps 17;			
QY	4	GFVTASGGQFVLNGLPYRGYGTNNYLSY-----QSHADVDVLAQAQAMNLSVIRTWGF	58
DB	30	GFVRTKGVQFSLNGYPYANGFNAYLWLVASDPQSORSKISTAFODASRHLTVARTWAF	89
QY	59	IDIGSLDSVPTIDCNKNGFFQYWDPTGAPAYNDGPTGLOGLDYAIASAAHGLRVIV	118
DB	90	SD-----GGYRALQYSPG-----SYNEDM--FQGLDFALAEARRHGKIL	128
QY	119	VLTDNWKFEFGMDQYDKW---YGLPY--HDNFTDPTQQAQYKKNVNLNRVNSITGYT	173
DB	129	SFANNYESFGGRQYVDNARSGRPVSSDEDDFTSLVKDFYKNIKAVLNRFNTFTKVH	188
QY	174	YKNDPTIFAWELANPRCVSGTSLPTSGTCTQATIVNVVDMSAYVKSIDPNHMYSVGDE	233
DB	189	YKDDPTIMAWELMNEPRCPD----PSGRAIQA----WITEMAAHVKSILDRNHLLEAGLE	240
QY	234	GFYIGSTOGSGWPNYNDPSD-GVDNNALLRVKNIDFGTYHLXPNYWGQNA-----DWCTQ	286
DB	241	GFY-GQSSPQSKTLNPPQGFQDFTANNRIPGIDFTVHSPDEWFPDSSQSQMDFLNK	299
QY	287	WIKDHIAAA-AIGKPTILEEF-----GWQTPDRDSVYOT-----WTQTVRTNGEAGW	333
DB	300	WLDIAHQDAQNVLHKPIILAEFGSKMKKPGYTPAORDIVFNTVYSKIYGSAKRGCAAAAGG	359
QY	334	NFWMLAGNVNGQYPNY-DGFNVYPPSSTATV	364
DB	360	LFWQLL--VNG--IDNFQDGYGIILSQSSSTV	387

RESULT 7
AAG42172
ID AAG42172 standard; Protein; 431 AA.
XX
AC AAG42172;
XX
DT 18-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 52560.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
OS Arabidopsis thaliana.
XX
PN EPI033405-A2.

XX	06-SEP-2000.	2000EP-0301439.
PD		
XX		
PF	25-FEB-2000;	99US-0121825.
XX		99US-012180.
PR	25-FEB-1999;	99US-0123180.
PR	05-MAR-1999;	99US-0123548.
PR	09-MAR-1999;	99US-0125788.
PR	23-MAR-1999;	99US-0125788.
PR	25-MAR-1999;	99US-0126264.
PR	29-MAR-1999;	99US-0126785.
PR	01-APR-1999;	99US-0127462.
PR	06-APR-1999;	99US-0128234.
PR	08-APR-1999;	99US-0128714.
PR	16-APR-1999;	99US-0129845.
PR	19-APR-1999;	99US-0130077.
PR	21-APR-1999;	99US-0130449.
PR	23-APR-1999;	99US-0130510.
PR	23-APR-1999;	99US-0130891.
PR	28-APR-1999;	99US-0131449.
PR	30-APR-1999;	99US-0132048.
PR	04-MAY-1999;	99US-0132407.
PR	05-MAY-1999;	99US-0132484.
PR	06-MAY-1999;	99US-0132485.
PR	06-MAY-1999;	99US-0132486.
PR	07-MAY-1999;	99US-0132487.
PR	11-MAY-1999;	99US-0132863.
PR	14-MAY-1999;	99US-0134256.
PR	14-MAY-1999;	99US-0134218.
PR	14-MAY-1999;	99US-0134219.
PR	14-MAY-1999;	99US-0134221.
PR	14-MAY-1999;	99US-0134370.
PR	18-MAY-1999;	99US-0134768.
PR	19-MAY-1999;	99US-0134941.
PR	20-MAY-1999;	99US-0135124.
PR	21-MAY-1999;	99US-0135353.
PR	24-MAY-1999;	99US-0135629.
PR	25-MAY-1999;	99US-0136021.
PR	27-MAY-1999;	99US-0136392.
PR	28-MAY-1999;	99US-0136782.
PR	01-JUN-1999;	99US-0137222.
PR	03-JUN-1999;	99US-0137528.
PR	04-JUN-1999;	99US-0137502.
PR	07-JUN-1999;	99US-0137724.
PR	08-JUN-1999;	99US-0138094.
PR	10-JUN-1999;	99US-0138540.
PR	10-JUN-1999;	99US-0138847.
PR	14-JUN-1999;	99US-0139119.
PR	16-JUN-1999;	99US-0139452.
PR	16-JUN-1999;	99US-0139453.
PR	17-JUN-1999;	99US-0139492.
PR	18-JUN-1999;	99US-0139454.
PR	18-JUN-1999;	99US-0139455.
PR	18-JUN-1999;	99US-0139456.
PR	18-JUN-1999;	99US-0139457.
PR	18-JUN-1999;	99US-0139458.
PR	18-JUN-1999;	99US-0139459.
PR	18-JUN-1999;	99US-0139460.
PR	18-JUN-1999;	99US-0139461.
PR	18-JUN-1999;	99US-0139462.
PR	18-JUN-1999;	99US-0139463.
PR	18-JUN-1999;	99US-0139750.
PR	18-JUN-1999;	99US-0139763.
PR	21-JUN-1999;	99US-0139817.
PR	22-JUN-1999;	99US-0139899.
PR	23-JUN-1999;	99US-0140353.
PR	23-JUN-1999;	99US-0140354.
PR	24-JUN-1999;	99US-0140695.
PR	28-JUN-1999;	99US-0140823.
PR	29-JUN-1999;	99US-0140991.
PR	30-JUN-1999;	99US-0141287.
PR	01-JUL-1999;	99US-0141842.
PR	01-JUL-1999;	99US-0142154.

PR	02-JUL-1999;	99US-0142055.
PR	06-JUL-1999;	99US-0142390.
PR	08-JUL-1999;	99US-0142803.
PR	09-JUL-1999;	99US-0142920.
PR	12-JUL-1999;	99US-0142977.
PR	13-JUL-1999;	99US-0143542.
PR	14-JUL-1999;	99US-0143624.
PR	15-JUL-1999;	99US-0144005.
PR	16-JUL-1999;	99US-0144085.
PR	19-JUL-1999;	99US-0144086.
PR	19-JUL-1999;	99US-0144325.
PR	19-JUL-1999;	99US-0144331.
PR	19-JUL-1999;	99US-0144332.
PR	19-JUL-1999;	99US-0144333.
PR	19-JUL-1999;	99US-0144334.
PR	19-JUL-1999;	99US-0144335.
PR	20-JUL-1999;	99US-0144352.
PR	20-JUL-1999;	99US-0144632.
PR	20-JUL-1999;	99US-0144684.
PR	21-JUL-1999;	99US-0144814.
PR	21-JUL-1999;	99US-0145086.
PR	21-JUL-1999;	99US-0145088.
PR	22-JUL-1999;	99US-0145085.
PR	22-JUL-1999;	99US-0145087.
PR	22-JUL-1999;	99US-0145089.
PR	22-JUL-1999;	99US-0145192.
PR	23-JUL-1999;	99US-0145145.
PR	23-JUL-1999;	99US-0145218.
PR	23-JUL-1999;	99US-0145224.
PR	26-JUL-1999;	99US-0145276.
PR	27-JUL-1999;	99US-0145913.
PR	27-JUL-1999;	99US-0145918.
PR	27-JUL-1999;	99US-0145919.
PR	28-JUL-1999;	99US-0145951.
PR	02-AUG-1999;	99US-0146386.
PR	02-AUG-1999;	99US-0146388.
PR	02-AUG-1999;	99US-0146389.
PR	03-AUG-1999;	99US-0147204.
PR	04-AUG-1999;	99US-0147302.
PR	05-AUG-1999;	99US-0147192.
PR	05-AUG-1999;	99US-0147260.
PR	06-AUG-1999;	99US-0147303.
PR	06-AUG-1999;	99US-0147416.
PR	09-AUG-1999;	99US-0147493.
PR	09-AUG-1999;	99US-0147935.
PR	10-AUG-1999;	99US-0148171.
PR	11-AUG-1999;	99US-0148319.
PR	12-AUG-1999;	99US-0148341.
PR	13-AUG-1999;	99US-0148565.
PR	13-AUG-1999;	99US-0148684.
PR	16-AUG-1999;	99US-0149368.
PR	17-AUG-1999;	99US-0149175.
PR	18-AUG-1999;	99US-0149426.
PR	20-AUG-1999;	99US-0149722.
PR	20-AUG-1999;	99US-0149723.
PR	20-AUG-1999;	99US-0149929.
PR	23-AUG-1999;	99US-0149902.
PR	23-AUG-1999;	99US-0149930.
PR	25-AUG-1999;	99US-0150566.
PR	26-AUG-1999;	99US-0150884.
PR	27-AUG-1999;	99US-0151065.
PR	27-AUG-1999;	99US-0151066.
PR	27-AUG-1999;	99US-0151080.
PR	30-AUG-1999;	99US-0151303.
PR	31-AUG-1999;	99US-0151438.
PR	01-SEP-1999;	99US-0151930.
PR	07-SEP-1999;	99US-0152363.
PR	10-SEP-1999;	99US-0153070.
PR	13-SEP-1999;	99US-0153758.
PR	16-SEP-1999;	99US-0154039.
PR	20-SEP-1999;	99US-0154779.
PR	02-JUL-1999;	99US-0155139.
PR	23-SEP-1999;	99US-0155486.
PR	24-SEP-1999;	99US-0155659.
PR	28-SEP-1999;	99US-0156458.
PR	29-SEP-1999;	99US-0156596.
PR	04-OCT-1999;	99US-0157117.
PR	05-OCT-1999;	99US-0157753.
PR	06-OCT-1999;	99US-0157865.
PR	07-OCT-1999;	99US-0158029.
PR	08-OCT-1999;	99US-0158232.
PR	12-OCT-1999;	99US-0158369.
PR	13-OCT-1999;	99US-0159293.
PR	13-OCT-1999;	99US-0159294.
PR	13-OCT-1999;	99US-0159295.
PR	14-OCT-1999;	99US-0159329.
PR	14-OCT-1999;	99US-0159330.
PR	14-OCT-1999;	99US-0159331.
PR	14-OCT-1999;	99US-0159637.
PR	14-OCT-1999;	99US-0159638.
PR	18-OCT-1999;	99US-0159584.
PR	21-OCT-1999;	99US-0160741.
PR	21-OCT-1999;	99US-0160767.
PR	21-OCT-1999;	99US-0160768.
PR	21-OCT-1999;	99US-0160770.
PR	21-OCT-1999;	99US-0160814.
PR	21-OCT-1999;	99US-0160815.
PR	22-OCT-1999;	99US-0160980.
PR	22-OCT-1999;	99US-0160981.
PR	22-OCT-1999;	99US-0160989.
PR	25-OCT-1999;	99US-0161404.
PR	25-OCT-1999;	99US-0161405.
PR	25-OCT-1999;	99US-0161406.
PR	26-OCT-1999;	99US-0161359.
PR	26-OCT-1999;	99US-0161360.
PR	26-OCT-1999;	99US-0161361.
PR	28-OCT-1999;	99US-0161920.
PR	28-OCT-1999;	99US-0161992.
PR	28-OCT-1999;	99US-0161993.
PR	29-OCT-1999;	99US-0162142.
Query Match 26.1%; Score 543.5; DB 21; Length 431;		
Best Local Similarity 35.5%; Pred. No. 1.le-36;		
Matches 139; Conservative 49; Mismatches 139; Indels 65; Gaps 17;		
Qy	4	GFVTASGCOFVLNGLPYRGCTNNYLSY-----QSHADVDVLAQAAMNLSVIRTWCF 58
Db	30	GFVRTKGVOFSLNGPYFYANGFNAYMLMYVASDPSQRSKISTAFQDASRHHGLTVARTWAF 89
Qy	59	IDIGSLDGSVPTIDGNKNGFYQYWDPSGTGAPAYNDGPTGLQGLDYATASAAHGLRVIV 118
Db	90	SD-----CGYRALQYSPG-----SYNEDM--FQGLDFALAEARRHGIKIL 128
Qy	119	VLTNDMKEFGGMDQYDKW---YGLPY--HDNFTDPRTOQAYKNVYNHLLNRVNSITGVT 173
Db	129	SFANNESFGGRKQYVDWARSGRPRVSSDEDDFTDSLKVDFYKNHIAKAVLNRNFTTKVH 188
Qy	174	YKNDPTIFAWELANEPRCVGSGCTLPTSTCTCTQATIVNVWDMSAYVKSIDPNHNVSVGDE 233
Db	189	YKDDPTIMAWELMEPRCPD-----PSGRATQA----WITEMAAHVKSIDRNHLLLEAGLE 240
Qy	234	GFYIGSTQSGWPNVPSD-GVDNNALLRVKNIDFGTVHLYPNYWGONA-----DMGTO 286
Db	241	GFY-GQSSPOSKTLNPPCOFCOTDFIANNRIPGIDFVTVHSPDEWFPDSSSEOSOMDFLNK 299
Qy	287	WKDHIANAA-AICKPTILEEF-----GMOTPDROSVYOT-----WTQVTRTNGEAGW 333
Db	300	WLDAHQDAQNVLHKPIIAEFKSGMKKPGYTPAQRODIVFNTVYSKIYGSAKRGGAAGG 359
Qy	334	NFWMLAGNVNCPYPNY-DGNFNVYPPSATV 364
Db	360	LFWOLL--VNG--IDNFODGYGIILSSOSTV 387

RESULT 8		
ABB93972	300	WLDHIQDAQNVLHKPIILAEFGKSMKKPGYTPAQRODIVFNTVYSKIYGSARKGAAAGG 359
ID ABB93972 standard; Protein; 431 AA.	334	NFMVLGNVNGOPYPNY-DGFNVYPPSTATV 364
XX		::
AC ABB93972;	360	LFWOLL--VNG--IDNFQDGYGIILSSQSSSTV 387
XX		
DT 31-MAY-2002 (first entry)		
XX		
DE Herbicidally active polypeptide SEQ ID NO 3183.		
XX		
KW Herbicidal; plant; agriculture; herbicide.		
XX		
OS Arabidopsis thaliana.		
XX		
PN WQ200210210-A2.		
XX		
PD 07-FEB-2002.		
XX		
PF 28-AUG-2001; 2001WO-EP09892.		
XX		
PR 28-AUG-2001; 2001WO-EP09892.		
XX		
PA (FARB) BAYER AG.		
XX		
PI Tietjen K, Weidler M;		
XX		
DR MPI; 2002-269010/31.		
XX		
PT Identifying plant target proteins for herbicidally active compounds,		
PT comprising aligning and comparing nucleic acid or amino acid sequences		
PT from plant with nucleic acid or amino acid sequences from non-plant		
PT organisms -		
XX		
PS Claim 5; SEQ ID NO 3183; 261pp + Sequence Listing; English.		
XX		
CC The invention relates to identifying target proteins		
CC (ABB90790-ABB94016) for herbicidally active compounds, comprising		
CC aligning and comparing nucleic acid or amino acid sequences from plant		
CC with nucleic acid or amino acid sequences from non-plant organisms using		
CC suitable search parameters, where plant sequences having an E-value		
CC greater by a factor of 3 than the E-value of most similar non-plant		
CC sequences are selected. The polypeptides or nucleic acids encoding them		
CC are useful for identifying modulators. The identified modulators are		
CC useful as herbicides.		
XX		
SQ Sequence 431 AA;		
Query Match 26.1%; Score 543.5; DB 23; Length 431;		
Best Local Similarity 35.5%; Pred. No. 11e-36;		
Matches 139; Conservative 49; Mismatches 139; Indels 65; Caps 17;		
QY 4 GFVTASGGQFVLNGLPYRYCGTNNYLSY-----QSHADVDDVLAKAQAAMNLSVIRTWGF 58		
DB 30 GFVRTKGQFSLNGPYPYANCENAYMLMYVASDPSQRSKISTAFODASRHGLTVARTWAF 89		
QY 59 IDIGSLDGSVPTIDGNKNGYFQWDPSTCAPAYNDGPTGLOGLDYAIASAAHGLRVIV 118		
DB 90 SD-----GGTRALQISPC-----SYNEDM--FQGLDFALAEARRHGKIIIL 128		
QY 119 VLFTNDMKEFGCMQYDKW---YGLPY--HDNFTYDPTQQAQKNVNHLLNRVNSITGVT 173		
DB 129 SFANNVESFCGRKQYVDWARSRRPVSSDEDDFTDSLVDKQYKNHIAKAVLNRRFTFKVH 188		
QY 174 KYNDPTIFANELANPRCVSGTLPTSGTCTQATIVNVQMSAYKYSIDPNHVMVSGDE 233		
DB 189 YKDPTIMAWELNPRCPSPD---PSGRAIOA---WITEMAAHVKSLDRNHLLEAGLE 240		
QY 234 GFYIGSTQCGSWPYNPSPD-GVDNNALLRVKNIDFCGYHLYPNYWGONA-----DWGQTQ 286		
DB 241 GFY-GOSSPOSKTLNPPGQGTDFIANNRIPGIDFVTVHSPDEWFPDSSEQSOMDFLNK 299		
QY 287 WIKDHIANAA-AIGKPTILEEF-----GWQTPDRDSVYQT-----WTQTVRTNGEAGW 333		
Db	300	WLDHIQDAQNVLHKPIILAEFGKSMKKPGYTPAQRODIVFNTVYSKIYGSARKGAAAGG 359
QY	334	NFMVLGNVNGOPYPNY-DGFNVYPPSTATV 364
Db	360	LFWOLL--VNG--IDNFQDGYGIILSSQSSSTV 387
RESULT 9		
AAG42171		
ID AAG42171 standard; Protein; 442 AA.		
XX		
AC AAG42171;		
XX		
DT 18-OCT-2000 (first entry)		
XX		
DE Arabidopsis thaliana protein fragment SEQ ID NO: 52559.		
XX		
KW Protein identification; signal transduction pathway; metabolic pathway;		
KW hybridisation assay; genetic mapping; gene expression control; promoter;		
KW termination sequence.		
XX		
OS Arabidopsis thaliana.		
XX		
PN EP1033405-A2.		
XX		
PD 06-SEP-2000.		
XX		
PF 25-FEB-2000; 2000EP-0301439.		
XX		
PR 25-FEB-1999; 99US-0121825.		
PR 05-MAR-1999; 99US-0123180.		
PR 09-MAR-1999; 99US-0123548.		
PR 23-MAR-1999; 99US-0125788.		
PR 25-MAR-1999; 99US-0126264.		
PR 23-MAR-1999; 99US-0126785.		
PR 01-APR-1999; 99US-0127462.		
PR 06-APR-1999; 99US-0128234.		
PR 08-APR-1999; 99US-0128714.		
PR 16-APR-1999; 99US-0129845.		
PR 19-APR-1999; 99US-0130077.		
PR 21-APR-1999; 99US-0130449.		
PR 23-APR-1999; 99US-0130510.		
PR 28-APR-1999; 99US-0130891.		
PR 30-APR-1999; 99US-0131449.		
PR 30-APR-1999; 99US-0132048.		
PR 04-MAY-1999; 99US-0132407.		
PR 05-MAY-1999; 99US-0132485.		
PR 06-MAY-1999; 99US-0132486.		
PR 06-MAY-1999; 99US-0132487.		
PR 07-MAY-1999; 99US-0132863.		
PR 11-MAY-1999; 99US-0134256.		
PR 14-MAY-1999; 99US-0134218.		
PR 14-MAY-1999; 99US-0134219.		
PR 14-MAY-1999; 99US-0134221.		
PR 14-MAY-1999; 99US-0134370.		
PR 18-MAY-1999; 99US-0134768.		
PR 19-MAY-1999; 99US-0134941.		
PR 20-MAY-1999; 99US-0135124.		
PR 21-MAY-1999; 99US-0135353.		
PR 24-MAY-1999; 99US-0135629.		
PR 25-MAY-1999; 99US-0136021.		
PR 27-MAY-1999; 99US-0136392.		
PR 28-MAY-1999; 99US-0136782.		
PR 01-JUN-1999; 99US-0137222.		
PR 03-JUN-1999; 99US-0137528.		
PR 04-JUN-1999; 99US-0137502.		
PR 07-JUN-1999; 99US-0137724.		
PR 08-JUN-1999; 99US-0138094.		
PR 10-JUN-1999; 99US-0138540.		
PR 10-JUN-1999; 99US-0138847.		
PR 14-JUN-1999; 99US-0139119.		
PR 16-JUN-1999; 99US-0139452.		


```
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140355.
PR 23-JUN-1999; 99US-0140356.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.

PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 26.1%; Score 543.5; DB 21; Length 442;
Best Local Similarity 35.5%; Pred. No. 1.2e-36;
Matches 139; Conservative 49; Mismatches 139; Indels 65; Gaps 17;

Oy 4 GFVTASGGQFVNLGLPYRYGGTNNYLSY-----OSHADVDVLLAKAQAMNLSVITWGF 58
Db 41 GFVRTKGVOFSLNGPYPYANGCFNAYMLMYVASDPQSORSKISTAFODASRHGLTVARTWAF 100
Oy 59 IDIGSLDGSVPTIDCNKNGFYFYQYWDPTGAPAYNDGPTGLOGLDYATASAAHGLRVIV 118
```


PR	26-OCT-1999;	99US-0161359.
PR	26-OCT-1999;	99US-0161360.
PR	26-OCT-1999;	99US-0161361.
PR	28-OCT-1999;	99US-0161920.
PR	28-OCT-1999;	99US-0161992.
PR	28-OCT-1999;	99US-0161993.
PR	29-OCT-1999;	99US-0162142.
Query Match 26.1% Score 543.5; DB 21; Length 443;		
Best Local Similarity 35.5% Pred. No. 1.2e-36;		
Matches 139; Conservative 49; Mismatches 139; Indels 65; Gaps		
QY	4	GFVTASGGQFVLNGLPYRYGCTNNYYSY-----QSHADVDVLAQAAMNLSVIRTWCF 58
DB	42	GFVRTKGVQFSLNGPYPYANGFNAYLWLMYVADSPORSKISTAFQDASRHGLTVARTWAF 101
QY	59	IDIGSLDGSVPTIDGNKNGFYFYQWDPSTGAPAYNDGPTGLOGLDYAIASAAAHGLRVIV 118
DB	102	SD-----GGYRALQYSPG-----SYNEDM--FQGLDFALAEARRHGKIKIL 140
QY	119	VLTDNWKFCGMDQYDKW---YGLPY--HDNFYTDPTQOAYKNWVHLLNRVNSITGVT 173
DB	141	SEANNYESFGGRKQYVDMAKSRGPRVSEDDFFTDLSLVKDYKKNHKAFLNRFNFTTKVH 200
QY	174	YKNDPTIFAWELANEPRCVSGTLPTSGTCTCTQATIVNVVDQMSAYVKSTDPNHMYSVGDE 233
DB	201	YKDDPTIMAWELMNEPRCPD-----PSGRAIOA----WITEMAAHVKSLLDRNHILLEAGLE 252
QY	234	GFYIGSTQSGWPNYDPSD-GVDNNALLRVKNIDFGTYVHLYPNYWGQNA-----DWGTO 286
DB	253	GFY-GQSSQSKTLNPPGQCTGDFIANNRIPGIDEVTVHSYPDEWFPSDESQSDMLNKK 311
QY	287	WIKDHIANAA-AICKPTILEEF-----GHQTPDRDSVYOT-----WTQTVRTNCEAGW 333
DB	312	WLDHAIODAQNVLHKPILIAEFSGMKMKPGYTPAQRDIVFNTVYSKIYGSARKGCAAGG 371
QY	334	NFWMLAGNVNCQYPNY-DGFNVYVPSSTATV 364
DB	372	LFWQLL--VNG--IDNFQDGYGIIILSSSSTV 399
RESULT 11		
ID	ABB92275	
XX	ABB92275 standard; Protein; 408 AA.	
AC	ABB92275;	
DT	31-MAY-2002 (first entry)	
DE	Herbicide; active polypeptide SEQ ID NO 1486.	
KW	Herbicide; plant; agriculture; herbicide.	
OS	Arabidopsis thaliana.	
PN	WO200210210-A2.	
XX	07-FEB-2002.	
PF	28-AUG-2001; 2001WO-EP09892.	
PR	28-AUG-2001; 2001WO-EP09892.	
PA	(FARB) BAYER AG.	
PI	Tietjen K, Weidler M;	
DR	WPI; 2002-269010/31.	
XX	Identifying plant target proteins for herbicidally active compounds,	
PT	comprising aligning and comparing nucleic acid or amino acid sequences	
PT	from plant with nucleic acid or amino acid sequences from non-plant	
PT	organisms -	

XX PS Claim 5: SEQ ID NO 1486; 261pp + Sequence Listing; English.

XX CC The invention relates to identifying target proteins

CC (AB990790-AB9904016) for herbicidally active compounds, comprising

CC aligning and comparing nucleic acid or amino acid sequences from plant

CC with nucleic acid or amino acid sequences from non-plant organisms using

CC suitable search parameters, where plant sequences having an E-value

CC greater by a factor of 3 than the E-value of most similar non-plant

CC sequences are selected. The polypeptides or nucleic acids encoding them

CC are useful for identifying modulators. The identified modulators are

CC useful as herbicides.

XX CC

XX SQ Sequence 408 AA:

Query Match 25.8%; Score 537; DB 23; Length 408;

Best Local Similarity 33.3%; Pred. No. 3.7e-36;

Matches 135; Conservative 74; Mismatches 128; Indels 68; Gaps 16;

QY 1 APA-GVTAAGGQFVNLGPPYRGYGGTNNYLSYQ-----SHADVDVLAQAQAMNLSVIR 54

DB 24 APSDGFVRNGVQFILNGKPPFYANGFNAYWLAYEATDPTTRFKITNVFQNAATSLGLTIAR 83

QY 55 TWGFDIGSLDGSVPTIDGNKNGFYQYWDPTGAPAYNDGPTGLGLOYATASAAHGL 114

DB 84 TWGFRO-GAIYRALQAPGS-----YDEQT-----FQGLDFVIAEAKRIGI 123

QY 115 RVIVLTLNDWKKEFGMDQYDKWYG-----LPYHDNFYDPTQOAYKNWNHLLNRVNSI 169

DB 124 KLILLVNNWDDYGGKKQYVDWARSKEVSSNDDFRNPVKDFYKNHVKTVLNRVNTF 183

QY 170 TGVYTKNDPTIFAMELANEPRCVGSLTPTSGTCTQATIVNVWDQMSAYVKSIDPNHMY 229

DB 184 TKVAYDEPAINAWQLMNEPRC-----GVDKSG-----KTLMDWINEMAFVKSVDPNHLLS 235

QY 230 VGDEGYIGST---QCGSWPYNDPSDGVNALLRVKNIDFGTYHLYPNYWGQADWGT- 285

DB 236 TGHEGYGSSPERKNSLNPVSANTVYGADFIANHNTDAIDFASMHCGSLWFORLDQNSR 295

QY 286 -----QWIKDHIANAAT-GKPTILEFGW--QTP-----DRDSVYOT-----WTQTVRT 327

DB 296 LAFIKRWLECHIEDAQNILKPKVILAEFGILGSDTPRYTLANRDGVFTTYYDIIVYSAQKG 355

QY 328 NCEAGNFWLGNVNGQYPYNDG---FNVIYPSSTATVLAESA 369

DB 356 GSAAGALFW-----EVISEGMSNFAGPSSTILSDKSSSTVNIISHA 396

RESULT 12

AAG79297

ID AAG79297 standard; Protein; 416 AA.

XX AC AAG79297;

XX X: 61-JUN-2002 (first entry)

XX DE Amino acid sequence of endo-beta-mannanase enzyme of coffee.

XX KW Endo-beta-mannanase; coffee; polysaccharide; mannan; endo-mannanase;

XX KW coffee bean; transgenic plant.

XX OS Coffea arabica.

XX PN EP1138771-A1.

XX PD 04-OCT-2001.

XX PF 30-MAR-2000; 2000EP-0201175.

XX PR 30-MAR-2000; 2000EP-0201175.

XX PA (NEST) SOC PROD NESTLE SA.

PI Marraccini P, Rogers J, Pridmore RD;

XX WPI: 2001-612542/71.

DR N-PSDB: AA165954.

XX PT New DNA encoding endo-mannanase of coffee, useful for hydrolysing

PT mannan polysaccharides in foods, cosmetics and pharmaceuticals ;

XX PS Claim 6: Page 11-13; 21pp; French.

XX CC The present sequence represents an endo-beta-mannanase enzyme of coffee.

CC The enzyme is involved in hydrolysis of polysaccharides that comprise

CC mannan molecules, linear or branched, linked together by beta(1-4) bonds.

CC The endo-mannanase is useful in foods, cosmetics and pharmaceuticals,

CC particularly for hydrolysis of mannan-containing polysaccharides. It is

CC also useful for synthesis and modification of such polysaccharides in

CC vitro. Treatment of coffee beans with mannanase increases the proportion

CC of dry matter extract and the extraction yield, while reducing the

CC quantity of sediment. Similar treatment of coffee liquor reduces the

CC amount of sediment produced by gelation of mannans. The endo-mannase

CC polynucleotide is useful for recombinant production of the enzyme, as

CC a source of probes and primers for in vivo detection and modification

CC of mannanase-encoding genes and for preparation of transgenic plants with

CC altered proportions of mannan-containing polysaccharides, especially

CC coffee beans, for modification of their aroma and structure.

XX CC

XX SQ Sequence 416 AA;

Query Match 25.6%; Score 532.5; DB 22; Length 416;

Best Local Similarity 32.3%; Pred. No. 8.9e-36;

Matches 134; Conservative 65; Mismatches 127; Indels 89; Gaps 19;

QY 5 FVTASGGQFVNLGPPYRGYGGTNNYLSY-----QSHADVDVLAQAQAMNLSVIRTWGFI 59

DB 44 FVKTRTEFVMGRPLYLNGFNAYWLAYMASDPSTRTKVSTTFOQASKYQCMNAARTWAFS 103

QY 60 DIGSLDGSVPTIDGNKNGFYQYWDPTGAPAYNDGPTGLGLOYATASAAHGLRIVV 119

DB 104 DGG-----YRALQSPG--SYNEDM--FKGLDFVYSEAKYGIHLILT 142

QY 120 LTNDKKEFGMDQYDKW-----YGLPYHDNFYDPTQOAYKNWNHLLNRVNSITGVY 174

DB 143 LVNWEYIGKKQYVQWAKDQGHYLNNDDDFFTDPIVRGYFKNHIKTVLTRINSITGLAY 202

QY 175 KNDPTIFAMELANEPRCVGSLTPTSGTCTQATIVNVWDQMSAYVKSIDPNHMYVSGDEG 234

DB 203 KODPTIFAMELANEPRCQSD-----LSGKAIO-----DWISEMATHVKSIDSDHLLDGLGEC 254

QY 235 FYIGSTQSGWGP-----YNDPSDGVNALLRVKNIDFGTYHLYPNYWGQADWGTQ 286

DB 255 FY-----GESVPQKKEYNPGYQVGTDFISNN---RIVQVDFATIHLYPDQWPNNDQTO 306

QY 287 -----WIKDHIANAAT-AIGKPTILEEF-----QWOTPDQSDS-----VYOT--WTQTVR 326

DB 307 AQFVDRWIKEHIDDSKYLLEKPLLLTEFGKSSRSPGYQVAKRDAYLSHIYDTIYACAATR 366

QY 327 TNGEAGWN-FWMLAGNVNGQYPYNY-----DGFNVYYP--STATVLAESAIAIST 374

DB 367 GGVCGCGNLFQWQMA-----PCMESWCGDYEVILEENPSTVGVIAQGSNRLSS 414

RESULT 13

ABB90809

ID ABB90809 standard; Protein; 431 AA.

XX AC ABB90809;

XX X: 31-MAY-2002 (first entry)

XX DE Herbicidally active polypeptide SEQ ID NO 20.

XX KW Herbicidal; plant; agriculture; herbicide.

```
OS Arabidopsis thaliana.
XX
PN WO200210210-A2.
XX
PD 07-FEB-2002.
XX
PF 28-AUG-2001; 2001WO-EP09892.
XX
PR 28-AUG-2001; 2001WO-EP09892.
XX
PA (FARB ) BAYER AG.
XX
PI Tietjen K, Weidler M;
XX
DR WPI; 2002-269010/31.
XX
XX
XX Identifying plant target proteins for herbicidally active compounds,
XX comprising aligning and comparing nucleic acid or amino acid sequences
XX from plant with nucleic acid or amino acid sequences from non-plant
XX organisms .
XX
XX Claim 5; SEQ ID NO 20; 261pp + Sequence Listing; English.
XX
XX The invention relates to identifying target proteins
XX (ABB90790-ABB94016) for herbicidally active compounds, comprising
XX aligning and comparing nucleic acid or amino acid sequences from plant
XX with nucleic acid or amino acid sequences from non-plant organisms using
XX suitable search parameters, where plant sequences having an E-value
XX greater by a factor of 3 than the E-value of most similar non-plant
XX sequences are selected. The polypeptides or nucleic acids encoding them
XX are useful for identifying modulators. The identified modulators are
XX useful as herbicides.
XX
XX Sequence 431 AA;
XX
XX Query Match 24.9%; Score 517.5; DB 23; Length 431;
XX Best Local Similarity 32.3%; Pred. No. 1.6e-34;
XX Matches 130; Conservative 69; Mismatches 132; Indels 71; Gaps 14;
XX
QY 4 GFVTASGGQFVNLGLPYRYGCTNNYLL-----SYQSHADVDVLAQAAMNLSVIRTW 56
DB 48 GFVGRNGTQFVNLGEQVYVNGFNAYWMTTAAADTASGRATVTTALRQASAVGMNVARIW 107
XX
QY 57 GFDIGSLDGSVPTIDGNKNGFYQYWDPTGAP-AYNDGPTGLOGLDYATASAAHGLR 115
DB 108 GF-----NEGDIYI---PLQISPGYSSEVY--FKGLDFVYVEAGRFNIK 145
XX
QY 116 VIVVLTNDWKEFGMDQYDKWYGLPYHDNFYDPTQOAYKNWVHLLNRVNSITGVYK 175
DB 146 LIISLVNFDYGGRRKKYVEWAGLDEPDEFYNSAVQFYKNHVKTVLTKRNTITGRMYK 205
XX
QY 176 NDPTIFAWELANPRCVSGTGLPTSGTCTQATIV-NWVDQMSAYVKSIDPNHMYSVGDEG 234
DB 206 DDPTIFSWELINEPRCNDSE-----TASNILQDWKEMASVYKSIDSNHLEIGLEG 256
XX
QY 235 FYICSTOGSGWPN-----DPSGVDNALLRVKNTDFGTYHLYPNW-----GON 280
DB 257 FY-GESIPERTVYPPGGRVLTGDTFTNN---QIPDIDFATIHYPDSWLPQSSRTGEQ 312
XX
QY 281 ADNGTOWIKDHIANA-AAIKPTILEEF-----GMDTPDRDSVYOTWTQVTRTNGEAG 332
DB 313 DTFVDRWIGAHIEDCDNIKKPLLITFEGSKSKYPGFSLEKRNFFORVYDVYIDSRAG 372
XX
QY 333 WN-----FWMLAGNVNQPNYDGFNVYPSSTATVLASEA 369
DB 373 GSCTGGVFQWLTNRTGLLGDGYEVFMQAGPNTTQALIADQS 414
XX
RESULT 14
ID ABB92274
XX ABB92274 standard; Protein: 414 AA.
XX
AC ABB92274;
```

```
XX
XX 31-MAY-2002 (first entry)
XX
XX Herbicidally active polypeptide SEQ ID NO 1485.
XX
XX Herbicidal; plant; agriculture; herbicide.
XX
XX Arabidopsis thaliana.
XX
XX WO200210210-A2.
XX
XX 07-FEB-2002.
XX
XX 28-AUG-2001; 2001WO-EP09892.
XX
XX 28-AUG-2001; 2001WO-EP09892.
XX
XX (FARB ) BAYER AG.
XX
XX Tietjen K, Weidler M;
XX
XX WPI; 2002-269010/31.
XX
XX Identifying plant target proteins for herbicidally active compounds,
XX comprising aligning and comparing nucleic acid or amino acid sequences
XX from plant with nucleic acid or amino acid sequences from non-plant
XX organisms .
XX
XX Claim 5; SEQ ID NO 1485; 261pp + Sequence Listing; English.
XX
XX The invention relates to identifying target proteins
XX (ABB90790-ABB94016) for herbicidally active compounds, comprising
XX aligning and comparing nucleic acid or amino acid sequences from plant
XX with nucleic acid or amino acid sequences from non-plant organisms using
XX suitable search parameters, where plant sequences having an E-value
XX greater by a factor of 3 than the E-value of most similar non-plant
XX sequences are selected. The polypeptides or nucleic acids encoding them
XX are useful for identifying modulators. The identified modulators are
XX useful as herbicides.
XX
XX Sequence 414 AA;
XX
XX Query Match 24.8%; Score 516; DB 23; Length 414;
XX Best Local Similarity 32.3%; Pred. No. 2.1e-34;
XX Matches 133; Conservative 59; Mismatches 130; Indels 90; Gaps 14;
XX
QY 4 GFVTASGGQFVNLGLPYRYGCTNNYLLSYQ-----SHADVDVLAQAAMNLSVIRTWGF 58
DB 30 GFVSRKGVQFVLNGKPFYANGFNAYWLAYEATDSTTRFKITYVFQONATIHDLTIVRTWGF 89
XX
QY 59 IDIGSLDGSVPTIDGNKNGFYQYWDPTGAPAYNDGPTGLOGLDYATASAAHGLRVIV 118
DB 90 RDGYRALQI-----APGVYDEKT-FOGLDFAIAEAKRLGKIMII 128
XX
QY 119 VLTNDWKEFGMDQYDKW-----YCLPYHDNFYDPTQOAYKNWVHLLNRVNSITGV 173
DB 129 TFVNNYSDFGGRKQYVDWAKNTQNVSSDDDFYTNPLVKQYVKNHVKTMVRNVTFTKVE 188
XX
QY 174 YKNDPTIFAWELANPRCVSGTGLPTSGTCTQATIVNWVDQMSAYVKSIDPNHMYSVGDE 233
DB 189 YKDEPTINGMELMNEPOCRAD---PSGKTLTA---WMNEMALYVKSVDKSHLLSTGLE 240
XX
QY 234 GFYICSTOGSGWPNYNDPS-----DGVNNDALLRVKNTDFGTYHLYPNWVGONADWGT 285
DB 241 GFY-----GDSPPORKTSLNPVAANVLGTDFIANKHKLDAIDFASIHYSYVDFWFFNLDEKS 295
XX
QY 286 -----QWIKDHIANA-AAI-GKPTILEEFGMDT-----PDRDSVYOTWTQVTRTNGE- 330
DB 296 RLNLLRWLECHLEDAQNILKKPLILGEFGKPTNTPGYTOAQRDAVFNATFDITIYESAEK 355
XX
QY 331 -----AGNFWMLAGNVNQPNYDGFNVYYP-----SSTATVLASEA 369
DB 356 GGPAAGALFWHVIS-----DCMNNFKDPLSVLSENSTTVNITTES 397
```


GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 13, 2002, 11:50:41 ; Search time 17.5108 Seconds
(without alignments)
2058.756 Million cell updates/sec

Title: US-09-917-378-3

Perfect score: 2079

Sequence: 1 APAGFVTASGGQFVLNGLPY.....YYPSTATVLASEALISTG 375

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 9613422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : PIR_73.*

1: pir1.*

2: pir2.*

3: pir3.*

4: pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	811	39.0	669	D72278	endo-1,4-beta-mann
2	517.5	24.9	411	D86153	hypothetical prote
3	498	24.0	448	T48214	endo-1,4-beta-mann
4	495	23.8	431	T09048	probable mannan en
5	426.5	20.5	369	T04323	mannan endo-1,4-be
6	411.5	19.8	403	A84592	(1-4)-beta-mannan
7	399.5	19.2	442	T87348	mannanase, probabl
8	171.5	8.2	589	T35784	probable secreted
9	143	6.9	2817	T97033	uncharacterized pr
10	140.5	6.8	649	AH3546	outer membrane pro
11	138.5	6.7	397	A35136	cellulase (EC 3.2.
12	136.5	6.6	722	T96986	endo-1,4-beta gluc
13	134	6.4	482	JC7332	endoglycosylcerami
14	134	6.4	1684	S10789	amylase A-180 - al
15	130.5	6.3	456	C42360	cellulase (EC 3.2.
16	128	6.2	747	B47093	cellulase (EC 3.2.
17	126	6.1	433	JH0158	cellulase (EC 3.2.
18	125.5	6.0	504	S54744	probable non-proce
19	125	6.0	425	C97013	beta-galactosidase
20	124.5	6.0	2228	T97942	beta-galactosidase
21	124.5	6.0	2233	T895075	beta-galactosidase
22	123.5	5.9	1155	T71456	conserved prolysin
23	122.5	5.9	597	T90481	probable pyrolysin
24	117.5	5.7	1144	A36968	PI-like adhesin pr
25	116	5.6	516	JE0134	mannan endo-1,4-be
26	114	5.5	430	S55325	endo-beta-1,6-gluc
27	112.5	5.4	566	A40589	cellulase (EC 3.2.
28	112	5.4	407	T39282	probable exo-beta-
29	112	5.4	644	A43370	capsid protein - N

RESULT 1

D72278

endo-1,4-beta-mannosidase - Thermotoga maritima (strain MSB8)

C:Species: Thermotoga maritima

C:Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 21-Jul-2000

C:Accession: D72278

R:Nelson, K.E.; Clayton, R.A.; Gill, S.R.; Gwinn, M.L.; Dodson, R.J.; Haft, D.H.; Hic

Garrett, M.M.; Stewart, A.M.; Collon, M.D.; Pratt, M.S.; Phillips, C.A.; Richardson,

C.M.

Nature 399, 323-329, 1999

A:Title: Evidence for lateral gene transfer between Archaea and Bacteria from genome

A:Reference number: A72200; MUID:99287316; PMID:10360571

A:Accession: D72278

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-669 <ARN>

A:Cross-references: GB:AE001779; GB:AE000512; NID:94981777; PIDN:AAD36302.1; PID:9498

A:Experimental source: strain MSB8

C:Genetics:

A:Gene: TM1227

Query Match 39.0%; Score 811; DB 2; Length 669;

Best Local Similarity 42.8%; Pred. No. 1-le-52;

Matches 169; Conservative 61; Mismatches 121; Indels 44; Gaps 13;

Qy 5 FVTASGGQFVLNGLPYRYGGTNNYLSYOSHADVDDVLAKAOMNLSVITWGFIDIGSL 64

Db 23 FVKVENGKFAALNGKFRFGISNNYMHYKSNRMIDSVLESARDMGIKVLRWGF----L 77

Qy 65 DGSVPTIDGNKNGHYFYQWDISTCAPAYNDG-----PTGLOGLDYAIASAAHGLKLVVI 120

Db 78 DGESYCRDN-----TYMHPGPGVGVGPEGISNAQSGFERLDYTVAKAKELGILKLVIV 131

Qy 121 TNDKKEFGCMQYDKWGLPYHDNFYDPTDPTQOAYKKNVHILLNRVNSTGTGYTNDPTI 180

Db 132 VNNHDDFGGMNQYVWFCTGTHDDFYRDEKIKKEYKKYVSLVNVHNTYTGVPYREEPTI 191

Qy 181 FAWELANPRCVGSGTLPTSCCTQATIVNNVDOMSAVYKSIDPNHMSVSGDEGF---YI 237

Db 192 MAWELANPRC-----ETDKSGN---TLVENVKENSIVKSLDPNULVAVGDEGFFSNYE 243

Qy 238 GSTGSG---WPYNDPSCVDNNALLRVKNIDFGTYHLYPNYWGON----ADWGTQWIKD 290

Db 244 GFKPYGGEAEAYNGWS-GVDMKLLSTETVDYDCTFHLYPHSHWGVSPENYAQWAKKIED 302

Qy 291 HIANAAAIKPTILEEFG--WQTP--DRDSVYQYTWTTVTRTNGEAGNFWMLAGNVNGOP- 346

Db 303 HIKAKELGKPVLEEYGIKAPAPVNTAIYRLNNDLVYDLGGGAMFWMLAGICEGSDR 362

Qy 347 -----YPNYDGFNYPYSSTATVLASE--ALAISTG 375

Db 363 DERGYYPDYDGFIRVNDSDSPEALIREYAKLFNTG 397

ALIGNMENTS

RESULT 2

D86153
hypothetical protein T6A9.1 - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cross)
C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Dec-2001
C:Accession: D86153
R:Theologos, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;
ansen, N.F.; Hughes, B.; Huizar, L.
Nature 408, 816-820, 2000
C:A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziani,
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shin, P.; Southwick, A.M.; Sun, H.; Tallon,
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A:Reference number: A86141; MUID:21016719; PMID:11130712
C:Accession: D86153
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-411 <STO>
A:Cross-references: GB:AE005172; NID:g9857528; PIDN:AAG00883.1; GSPDB:GN00141
C:Genetics:
A:Map position: 1

Query Match 24.9%; Score 517.5; DB 2; Length 411;
Best Local Similarity 32.3%; Pred. No. 4.5e-31;
Matches 130; Conservative 69; Mismatches 132; Indels 71; Gaps 14;
QY 4 FVTASGGQFVLNGLPYRYGGTNNYLL-----SYQSHADVDVLAQAQANLNVIRTW 56
DB 28 GFVGRNGTQFVLNGEYVYVNGFNAYMMTAAADTASGRATVTTALRQASAVGHWARIW 87
QY 57 GFIDIGSLDGSVPTIDGNKNGCFYQYWDPSGAP-AYNDGPTGLOGLDYAIASAAHCLR 115
DB 88 GF-----NEGDYI-----PLQISPCSYSEDV--FKGLDFVYVEAGRENK 125
QY 116 VIVVLTNDWKEFGGMDQYDKWGLPYHDNFYDPTQOAYKNWVHLLNRVNSITGVTYK 175
DB 126 LILSIVNWFEDYGRKKYVEMAGLDEPFTYNSAVKQFYKNHVKTVLTAKNTITGRMYK 185
QY 176 NDPTIFAWELANPRCVGSGTLPTSGTCTQATIVNNVDMSAYVKSIDPNHIMVSGDEG 234
DB 186 DDPTIFSWELLNEPRCNDSE-----TASNILQDWKEMASVKSIDSNNHLEIGLEG 236
QY 235 FYTGSTGCGSWPYN-----DPSGVDNNALLRVKNIIDFGTYHLYPNWY-----GQN 280
DB 237 FY-GESTIPERTVYNPGGRVITGTDFITNN---QIPDIDFATIHYPDSMLPLOSSRTGEQ 292
QY 281 ADMGTQWIKDHIANA-AAIGKPTILEEF-----GWOTPDSDSVYQYTWQTVHTNGEAG 332
DB 293 DTFVDRIWIGAHIEDCDNIHKPLLITFEFGKSSYPGHSLEKRNKFFORVYDVYDSARAG 352
QY 333 WN-----FWMLAGNVNGQPYNDGFNFVYPSSTATVLASEA 369
DB 353 GSCTGGVFMOLTNRITGLLDGVEVFMAGPNTTAQIIADQS 394

RESULT 3

T48214
endo-1,4-beta-mannosidase-like protein - Arabidopsis thaliana
N:Alternate names: protein T20L15.200
C:Species: Arabidopsis thaliana (mouse-ear cross)
C:Date: 20-Apr-2000 #sequence_revision 20-Apr-2000 #text_change 20-Apr-2000
C:Accession: T48214
R:Bevan, M.; Peters, S.A.; van Staveren, M.; Dirkse, W.; Stiekema, W.; Bancroft, I.; Mey
submitted to the Protein Sequence Database, March 2000
A:Reference number: 224488
A:Accession: T48214
A:Status: preliminary
A:Molecule type: DNA

A:Residues: 1-448 <BEV>
A:Cross-references: EMBL:AL162351
A:Experimental source: cultivar Columbia; BAC clone T20L15
C:Genetics:
A:Map position: 5
A:Introns: 34/2; 126/3; 191/3; 303/2
A:Note: T20L15.200

Query Match 24.0%; Score 498; DB 2; Length 448;
Best Local Similarity 32.2%; Pred. No. 1.4e-29;
Matches 132; Conservative 66; Mismatches 130; Indels 82; Gaps 17;
QY 6 VTASGGQFVLNGLPYRYGGTNNYLL-----SYQSHADVDVLAQAQANLNVIRTWGFIID 60
DB 48 VORKGMQFTLNGQPFYVNGFNMTLAADNSTRGKVKTEVPQASAVGMTVGRWAFND 107
QY 61 IGSIDGSVPTIDGNKNGCFYQYWDPSGAPAYNDGPTGLOGLDYAIASAAHCLRIVVL 120
DB 108 -----GQNRALQKSPSVYDEEV-FKALDFVLSKARKYKIKLILSL 146
QY 121 TNDWKEFGGMDQYDKW-----YGLPYHDNFYDPTQOAYKNWVHLLNRVNSITGVTYK 175
DB 147 VNNMDAYGGKAQYVYKMGNASGLNLTSDDDFTNPTLRFYQSHVTVLNRVNTFTNITYK 206
QY 176 NDPTIFAWELANPRCVGSGTLPTSGTCTQATIVNNVDMSAYVKSIDPNHIMVSGDEGF 235
DB 207 NDPTIFAWELMNEPRCPSD-----PSGDKLQ-----SWIQEMAVFVKSIDAKHLVLEIGLEGF 258
QY 236 YIGS-----TGSGWPY---NDPSGVDNNALLRVKNIIDFGTYHLYPNWYNGONA-----DW 283
DB 259 YGPSAPARTFRPNPYPAAQVCTDFIRNNQVL---GIDFASVHVYPDSWISPAVSNSPILEF 315
QY 284 GTQWIKDHIANA-AA-IGKPTILEEF-----GWOTPDSD-----SVYQYTWQTVHTNGEAG 331
DB 316 TSSWQMAHVEDAEAMYLGMPLFTFEFGVSNHDPGFNTSFRDMLMTVYVYKMTLNSTRKGG-A 374
QY 332 GWNEWMLAGNVNGQPYN-----YDGFENVY---YPSSTATVLASEAALAI 372
DB 375 G-----AGSLVWQVFPQGAFFMDGQYAVYITRAITASKIISLOSRLAI 418

RESULT 4

T09048
probable mannan endo-1,4-beta-mannosidase (EC 3.2.1.78) - Arabidopsis thaliana
N:Alternate names: (1-4)-beta-mannan endohydrolase; protein F26K10.200
C:Species: Arabidopsis thaliana (mouse-ear cross)
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C:Accession: T09048
R:Bevan, M.; Lennard, N.; Quail, M.; Harris, B.; Rajandream, M.A.; Harrell, B.G.; Ban
submitted to the Protein Sequence Database, June 1999
A:Reference number: 216533
A:Accession: T09048
A:Molecule type: DNA
A:Residues: 1-431 <BEV>
A:Cross-references: EMBL:AL049803; ATSP:F26K10.200; GSPDB:GN000062
A:Experimental source: cultivar Columbia; BAC clone F26K10
C:Genetics:
A:Gene: ATSP:F26K10.200
A:Map position: 4
A:Introns: 120/3; 186/3; 227/3; 298/2
C:Keywords: glycosidase; hydrolase

Query Match 23.8%; Score 495; DB 2; Length 431;
Best Local Similarity 32.9%; Pred. No. 2.2e-29;
Matches 124; Conservative 55; Mismatches 124; Indels 74; Gaps 16;
QY 1 APAGFVTASGGQFVLNGLPYRYGGTNNYLSYQSHAD-----VDDVLAKAQAQANLNVSI 53
DB 37 AKLGFVVRNGTQFVDDKPLVYVGNWSYW--FMDHAYVDEHSRLVNGEMLEAGAKMGLTVC 94
QY 54 RTWGFIDIGSLDGSVPTIDGNKNGCFYQYWDPSGAPAYNDGPTGLOGLDYATASAAHIG 113
DB 95 RTWAF-----NDGYNALQI-----SPCRFDERV-FOALDHIATAEARKHD 133

C:Genetics:

A:Gene: CC0801

Query Match 19.2%; Score 399.5; DB 2: Length 442;

Best Local Similarity 26.8%; Pred. No. 3.1e-22;

Matches 118; Conservative 65; Mismatches 159; Indels 99; Gaps 16;

QY 2 PAGFVTASGQFVNLGILPHYGGTNNYLSY-----QSHADVDDV---LAKAQAANLSVIR 54
DB 27 PKGFVTVKDGRLSLDGCKPKYRFAGTNNWYAALCAPAGYDGLRLRRELDRLKAMGVNLR 86
QY 55 TWGFDIGSLDGSVPTIDGNKNGFYFOYWDPTGAPAYNDGPTGLGGLDYAIAASAAHGL 114
DB 87 I-----LGAGEQSPAKV-----AMDPTFRGPCEDYNADLLKGLDVTLAEMAQRDM 131
QY 115 RVIVLTLNDMKFCGMDYDKWYG-----LPYHDFNYDPRTOQAYKN 157
DB 132 KAVIYVNNFNDWSCGMPAYLNNWCDGPFQGDHPAYPHIYOYADYSARFYANQKANALFRH 191
QY 158 WVNHLNLRVNSITGVYTKNDPTIFAMELANEPRKCVSGTLPTSGTCTQATIVNNVDMSA 217
DB 192 YVTSLSVRSTVTKPKYRDPTINSQLANEPRPGGSDAF---GQSNMPAYQAWIRDTAG 248
QY 218 YVKSIDPNNHMSVGDGEFYIGSTGSCWHPYNDSPDGVNDNALLRVKNIDFGTYHLIPNYW 277
DB 249 LIRKLNQHLVSTGSG-----TMGC---MGLESCVIDAHIA---PPVIDYMTLHIWPNNW 297
QY 278 G-----QNAWDGQTQ---WTKDHIAANAAIGKPTILEEFCW-----OTPD 314
DB 298 GWISMTDOPSTYAGKQKCHDYVAOHIALAKRLNKLPTIEFGLIRDGROFTPGSPPTTYR 357
QY 315 DSVYQRTWQ-----IVYTING-EAGWNFWMLAGN-----VNGQPYPNYDG 352
DB 358 DRYKYTKMLDALADMKAGGPTAGVNFWMNGEGRAGQPDAPWFKKDKSYGVGDPPEQGL 417
QY 353 FNVY-YPSSTATVLASEALAI 372
DB 418 FGVFDADASTLAVIKAHAAV 438

RESULT 8

T35784

probable secreted protein - Streptomyces coelicolor

C:Species: Streptomyces coelicolor

C:Date: 05-Nov-1999 #sequence_revision 05-Nov-1999 #text_change 05-Nov-1999

C:Accession: T35784

R:Seeger, K.J.; Harris, D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A.

submitted to the EMBL Data Library, July 1998

A:Reference number: 221570

A:Accession: T35784

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-589 <SEE>

A:Cross-references: EMBL:AL031013; PIDN:CAA19789.1; GSPDB:SC08A6.17

A:Experimental source: strain A3(2)

C:Genetics:

A:Gene: SC08A6.17

Query Match

Best Local Similarity 8.2%; Score 171.5; DB 2: Length 589;

Matches 84; Conservative 38; Mismatches 108; Indels 131; Gaps 21;

QY 1 APAGFVTASGG-----QFVLNGLPYRYGGTNNYLSY-QSHADVDDVLAKAQAANLSVIR 55
DB 41 AAGSVVKVTCGGQDWLTVDGSPYTVKG-----LTWGPSADADRYLPLDASMGVNTIRT 95
QY 56 WCFIDIGSLDGSVPTIDGNKNGFYFOYWDPTGAPAYNDGPTGLGGLDYAIAASAAHGLR 115
DB 96 W-----GTDASSRPLID-----SAAAHGVK 115
QY 116 VIVLTLNDMKFCGMDYDKWYGLPYHDFNYDPRTOQAYKN-----WVNHLNLRV 167
DB 116 VI---AGFWLQPGGP-----GSGGCVNYLTD---TAYKQMLAEFRPWQOE----- 156

QY 168 SITGVYTKNDPTIFAMELANE-----PRCVGSLTLPITSGTCTQATIVNNVDMSAYVKS 222
DB 157 -----YKDHPPGVLMNMGVNESVLGLQNCYGGDELEHQ-----KDAYTTFVNDVAKKIIIV 206
QY 223 DPNHMSVGDGEFYIGSTGSCWHPY---NDI-SIKVDNNALLRVKNIDFGTYHLIPNYWGO 279
DB 207 DPNHPVTSTDA--WVGA-----WYVYOKNAP-----DLDLVAV--NSYDA 242
QY 280 NADGWOWIKDHTANAAAIKPTILEEFG-----WOTP-DKDSVYQRTWOTVTRNGEA-GW 333
DB 243 VCDVRSAAW-----EOGGYTKPYIVTESGPAGEVPPDDANGVPEEPTDRAKAOGYTDAW 296
QY 334 N 334
DB 297 N 297

RESULT 9

B97033

uncharacterized protein, related to enterotoxins of other Clostridiales [imported]

C:Species: Clostridium acetobutylicum

C:Date: 14-Sep-2001 #sequence_revision 14-Sep-2001 #text_change 14-Sep-2001

C:Accession: B97033

R:Nolling, J.; Breton, G.; Omelchenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.;

J.; Daily, M.J.; Bennett, G.N.; Koonin, E.V.; Smith, D.R.

J. Bacteriol. 183, 4823-4838, 2001

A:Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium

A:Reference number: A96900; MUID:21359325; PMID:21359325

A:Accession: B97033

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-2817 <KUR>

A:Cross-references: GB:AE001437; PIDN:AAK79053.1; PID:g15023993; GSPDB:GN00168

A:Experimental source: Clostridium acetobutylicum ATCC824

C:Genetics:

A:Gene: CAC1079

Query Match

Best Local Similarity 6.9%; Score 143; DB 2: Length 2817;

Matches 103; Conservative 57; Mismatches 162; Indels 132; Gaps 29;

QY 4 GFVTASGGQFVLN-----GLPYRYGGTNNY-----LSYQSHADVDDVIAK 44
DB 1530 GFVTLSGNTYFDSYCEMRIGLTYI---NNNYFFNSKGMETGWSLYRYANPNGLILT 1586
QY 45 A-QAMNLSVIRTWGFDIGSLDGSVPTIDGNKNGF-----YFOYWDPTGAPAY----- 92
DB 1587 GFQTIIN---GKTYFFNSDGLDYDIQYINGSYGFQDKNGVMLYCLQITGNTYVLSNGL 1643
QY 93 -NDGPTGLGGLDYAIAASAAHGLRV-LVLTNDWKKEFG----- 128
DB 1644 SQSGFITLNGKTYFDS--YVGNRTGLIONNNNYFFGDNMGTLQGWISODNLRYYANSS 1701
QY 129 -----GNDQYDKWYGLPYHDNFYTDPRTOQAYKNWVHLLNLRVNSITGVYTKNDPT-IFA 182
DB 1702 GVCLTGLQITID---GKKYFFNSVARMETGLVYIN-----NTYYG--FDNDGTLIYS 1747
QY 183 WELAN-EPRCVGS-GTLPTSGTCTQATIVNNVDQM--SAYVKSID---PNHMSVGDGEF 235
DB 1748 WHNINRGMYCFNTDGTGVT-----GWINYLGRSCYLDSSQGFSLTGLTLTIGNIY 1797
QY 236 YIGS--TOGSGWPNDSQGVNDNA---LLRVKNIDFGTYHLIPNYWGNADNGTOWIKD 290
DB 1798 YFGSDYSMTGWVTSQSSKYYFNESCIMLTGFTIDGNTYV-FDSY--GNSTTGTRSLN- 1853
QY 291 HIANAAAIKPTILEEFGWQTPDRDSVY--QWTQTVTRTNGEAGWFWMLAG----- 340
DB 1854 --GNCYGFNDGDMILT-GWOTISGNNYFENPDCTAKIGLNTYEGKTYYPSTGCGYTGTGII 1910
QY 341 NVNQDYP-NYDGF-----NVYPSSTATV 364
DB 1911 NINSNTYFCYDCALKTGWIRNNYIYYADNNGII 1944

Db 248 --ALNKNIALFATEWGTSDCNCGSGGTYYEESQRWVDFM-TQNKISWTNWVL 295

endoglycosylceramidase (EC 3.2.1.123) - *Rhodococcus* sp.
N:Alternate names: Cer glycanase; endoglycoceramidase
C:Species: *Rhodococcus* sp.
C:Date: 08-Sep-2000 #sequence_revision 08-Sep-2000 #text_change 08-Sep-2000
C:Accession: JC7332
R:Sakaguchi, K.; Okino, N.; Sueyoshi, N.; Izu, H.; Ito, M.
J. Biochem. 128, 145-152, 2000
A:Title: Cloning and expression of gene encoding a novel endoglycoceramidase of *Rhodococcus*
A:Reference number: JC7332
A:Accession: JC7332
A:Molecule type: DNA
A:Molecules: 1-482 <SAK>
A:Experimental source: strain C9
C:Comment: This enzyme, a member of type II enzyme, cleaves the glycosidic linkage between
C:Genetics:

QY 4 GEV'TASGGQF---VLNGLPYRYGCTNYYI,---SYQSHADVDVDVLAKAQAMNLSVIRI" 55

Db 31 GSVEVKQGEYRELNLQLENKPFSDNANVYEVLTDRPVGNGTSDNDSYGRPQ-----IDA #4

Qy	56	WG	-----	FIDIGS	-----	LIDGSVPTIDGNKNGPFFQY	82
						:	

DB	85	WGYNIGTFHGGIRGLTRKLECYFTDLGTNAIWISAPWEQMHGV-----GCKDGDFAHY	140
QY	83	WDPSTGAPAYNDGGTPTGLGSLDYA-----[ASAAAHGLRVI]---VVLTN---	124

```

Db      141  -----GYHGYYGLDFTAMQNMCTIDEMREFVDLAHSLGTRVVDIDVMNHVGY 188
          | | | | | : | : | | | : | :

```

Qy 123 ----DWKEFGGMDQYDKWYGLPYHDNFYTDPRTOQAY-KNWNH--LLKNVNSJTCGVYK 175

DB 189 PTIVDMHEFG---FGDTGGGLP-----RDWTPNQAGQGNWIIITNDIMNKDNEAAWNNW 238

QV 176 NDPTIFAWELANFPRCVGS-----CTLP-----TSCCTCTGTAATLVN-LVDNMSAVVKS 222

239 GSDWI RADETAGYDNCGGSEQTMCIGFLPDI KTEVTTTCVDLPPIILRNKNDOASGY ---- 294

QY 223 DPHMVSVDGEGFYIGSTOGSGWPYNDRPSDGVDDNALLRVKNIDFCTYHLYPNWGCNAD 282
| : : : | : : : | : :

DB 295 -----EDWFWPAE-----PYRQDLNIAPKDYLIK..... 319

OV 283 WCTGOWIKWHIANAATIT--CKPDIIFEECWOTDNDSDSVUOYUOMVDMNCFACIL--MUNAM 290

[illegible]

Qy 339 A 339 —

Db 380 A 380

RESULT 15
C42360

cellulase (EC 3.2.1.4) E5 precursor - Thermomonospora fusca
N:Alternate names: endo-1,4-beta-glucanase
C:Species: Thermomonospora fusca

C:Date: 24-Jul-1992 #sequence_revision 24-Jul-1992 #text_change 20-Mar-1998
C:Accession: C42360

A>Title: DNA sequences of three beta-1,4-endoglucanase genes from *Thermomonospora fus*

F:38-137/Domain: bacterial cellulose-binding domain homology <BCB>
Query Match 6.3%; Score 130.5; DB 2; Length 466;
Best Local Similarity 22.3%; Pred. No. 0.037;
Matches 82; Conservative 46; Mismatches 116; Indels 123; Gaps 21;
QY 16 NGIIPRYGCTNNYYLSYQSHADVD---DVIA---KQAMNLSVIRTWGFIDJGSLDGSVP 69
DB 182 HGNPVQLRCMSTHG IQWFDHCLTDSLDALAYDWKADIIRLSM----- 224
QY 70 TIDGNKNGFYFQWDPSTGAPYNDGPTGL-QGLDYAIAASAAHGLRVIVVLTNDWKEFG 128
DB 225 -----YIQ-----EDGYETNPRGFTDRIDQLIDMATARGLYVIV---DWHI-- 262
QY 129 GMDQYDKWGLPYHDNFYDPRTOQAYKNWVHILLNRVNSITGTYKNDPTIFAWELANE 188
DB 263 -LTPGDPHYNLDRAKTFFAEIAORHASK-----TNVLY-----ETANE 299
QY 189 PRCVGSGTLPTSGTCTQATIVNMVDQMSAYVKSIDPNHMVSVGDEGF-YIGSTQSGWPY 247
DB 300 PNCV-----SWASIKSYAEVIVPIRORPDSVLIIVGTRGSSSLGVSEGS--- 345
QY 248 NDPSCGVDDNALLRVKNIDFGTYHLYPNYWGONADWGTOMIKDHIANA---AAIGKPTIL 304
DB 346 --PAETIANP--VNASNIMY-AFHFY-----AASHRDNYLNALKEASELFPVFV 389
QY 305 EEFGWQTPDRDSV--YQTWTQTVRTNCE--ACHNFWMLAGNVNQCYPNY-DGFN---VY 356
DB 390 TEFGTETTYTDCGANDFOMADRYIDLMAERKIGWTKW-----NYSDDFRSGAVF 437
QY 357 YPSSTAT 363
DB 438 QPCTCAS 444

Search completed: November 13, 2002, 11:57:33
Job time : 19.5108 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 13, 2002, 11:47:41 : Search time 8.62069 Seconds
(without alignments)
1804.220 Million cell updates/sec

Title: us-09-917-378-3

Perfect score: 2079

Sequence: 1 APAGFVTAGCGFVLNGLPY.....YPSSTATVLASEALAISTG 375

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_40.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	179.5	8.6	562	1	GUN1_ACICE	P54583 acidothermu
2	138.5	6.7	397	1	GUN_PAEP0	P23548 paenibacill
3	128	6.2	747	1	GUND_CELFI	P50400 cellulomona
4	127	6.1	459	1	GUNA_STRLI	P27035 streptomyc
5	126	6.1	484	1	GUNA_XANCP	P19487 xanthomonas
6	125.5	6.0	504	1	GUNW_ERWCA	O59395 erwinia car
7	123.5	5.9	466	1	GUN5_THERFU	Q01786 thermomonos
8	115.5	5.6	879	1	MANB_CAPHI	O95327 capra hircu
9	113	5.4	1953	1	BIGA_SALTY	P25927 salmonella
10	112.5	5.4	444	1	GUNN_ERWCA	O59394 erwinia car
11	112.5	5.4	566	1	GUNC_CLOTM	O05332 clostridium
12	110.5	5.3	505	1	GUNV_ERWCA	O47096 erwinia car
13	110.5	5.3	825	1	GUN3_BACSA	P19570 bacillus sp
14	110	5.3	563	1	GUNE_CLOTM	P04956 clostridium
15	109.5	5.3	383	1	MANA_STRLI	P51529 streptomyc
16	109	5.2	448	1	GUN_CLOSA	P15704 clostridium
17	107.5	5.2	426	1	GUNZ_ERWCH	P07103 erwinia chr
18	107	5.1	409	1	GUN2_BACSA	P06565 bacillus sp
19	106.5	5.1	879	1	MANB_BOVIN	Q29444 bos taurus
20	106.5	5.1	1742	1	GUNA_CALSA	P22534 caldocellum
21	106	5.1	400	1	GUN5_BACN	O85465 bacillus ag
22	104	5.0	598	1	BGAL_XANMG	P48982 xanthomonas
23	104	5.0	603	1	BGLR_ECOLI	P05804 escherichia
24	104	5.0	900	1	GUNH_CLOTM	P16218 clostridium
25	103.5	5.0	426	1	GUN2_RALSO	P17974 raistonia s
26	103.5	5.0	1039	1	GUNE_CALSA	P10474 c endogluc
27	102.5	4.9	651	1	BGLR_HUMAN	P08236 homo sapien
28	102.5	4.9	991	1	ENV_VILV2	P23423 visna lenti
29	102	4.9	486	1	YAIT_ECOLI	P77199 escherichia
30	102	4.9	658	1	GUN3_FIBSU	P14250 fibrobacter
31	101.5	4.9	1365	1	GTF5_STRDO	P07983 streptococc
32	99.5	4.8	499	1	GUN1_BACSA	P10475 bacillus su
33	99.5	4.8	499	1	GUN2_BACSA	

ALIGNMENTS

RESULT 1

ID	GUN1_ACICE	STANDARD	PKT	562 AA
AC	P54583			
DT	01-OCT-1996 (Rel. 34, Created)			
DT	01-OCT-1996 (Rel. 34, Last sequence update)			
DT	01-OCT-1996 (Rel. 34, Last annotation update)			
DE	Endoglucanase E1 precursor (EC 3.2.1.4) (Endo-1,4-beta-glucanase E1)			
DE	(Cellulase E1) (Endocellulase E1).			
OS	Acidothermus cellulolyticus.			
OC	Bacteria; Actinobacteria; Actinobacteriia (class); Actinobacteridae;			
OC	Actinomycetales; Frankineae; Acidothermaceae; Acidothermus.			
OX	NCBI_TaxID=28049;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=ATCC 43068 / 11b;			
RA	Laymon R.A., Himmel M.E., Thomas S.R.;			
RL	Submitted (AUG-1995) to the EMBL/GenBank/DDBJ databases.			
RN	[2]			
RP	X-RAY CRYSTALLOGRAPHY (2.4 ANGSTROMS) OF 42-398.			
RX	MEDLINE=96346058; PubMed=8719854;			
RA	Sakon J., Adney W.S., Himmel M.E., Thomas S.R., Kaplun P.A.;			
RT	"Crystal structure of thermostable family 5 endocellulase E1 from			
RT	Acidothermus cellulolyticus in complex with cellobiose.;"			
RL	Biochemistry 35:10648-10660(1996).			
CC	-I- FUNCTION: THERMOSTABLE ENZYME WITH AN OPTIMAL TEMPERATURE OF 81			
CC	DEGREES CELSIUS. HAS A VERY HIGH SPECIFIC ACTIVITY ON			
CC	CARBOXYMETHYLCELLULOSE.			
CC	-I- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic			
CC	linkages in cellulose.			
CC	-I- SIMILARITY: BELONGS TO CELLULOSE FAMILY A (FAMILY 5 OF GLYCOSYL			
CC	HYDROLASES).			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed usage by and for commercial			
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/			
CC	or send an email to license@isb-sib.ch).			
CC	-----			
EMBL	U33212; AAA75477.1; -			
DR	PDB; 1ECE; 14-OCT-96.			
DR	InterPro: IPR001919; Bac_cellose-bind.			
DR	InterPro: IPR001547; GH_5.			
DR	Pfam: PF00150; cellulase; 1.			
DR	Pfam: PF00553; CBM_2; 1.			
DR	PROSITE: PS00659; GLYCOSYL-HYDROL_F5; 1.			
KW	Cellulose degradation; Hydrolase; Glycosidase; Signal; 3D-structure.			
FT	SIGNAL 1 41			
FT	CHAIN 42 562			
FT	DOMAIN 42 400			
FT	DOMAIN 401 461			
FT	DOMAIN 462 562			
FT	ACT_SITE 203 203			
FT	ACT_SITE 323 323			
FT	NUCLEOPHILE.			

```
FT DISULFID 75 161
FT DISULFID 209 212
SQ SEQUENCE 562 AA; 60747 MW; 84E6256406A35041 CRC64;

Query Match
Best Local Similarity 8.6%; Score 179.5; DB 1; Length 562;
Matches 92; Conservative 50; Mismatches 129; Indels 113; Gaps 24;

QY 1 APAGFTASGQGV-LNGLPRYCGTN-----NXYLSYQSHADVDVLLAKAQAAMNLSV 52
DB 42 AGCGYHTSREILDANNVPRVRIAGINWFGFTCNVYVHGLWSKDYKSMQDIKSLGYNT 101
QY 53 IRTWGFIDIGSLGDSVPTIDGNKNGFYQYWDPTGAPAYNDGPTGLQGLDYAIAASAAH 112
DB 102 IRL-PSYSDILKPGCTMP-----NSINFQNMQD-----LOGLSLOVMDKIVAYAGQI 148
QY 113 GLRVIVLVNDWKEFGMDYQKWKYGLPYHDNFYDPTQOAYKNVHLLNRVNSITGV 172
DB 149 GLRII--LDRIHRPDCSG--QSALWY-----TSSVSEATWISDL-----QALAQ 187
QY 173 TYKNDPTIFAMELANEPR---CVSGTLPSTGCTQATIVNW---VDMSAYVKSIDPNH 226
DB 188 RYKGNPTVCGFDLNEPHDPACWGCD-PS-----IDWRLAERACNAVLSVNPNL 237
QY 227 MVSU-----GDEFGYICSTGSG-WPYNDPSGDVNDNALLRVKN-----IDFGTYHL 272
DB 238 LIFVEGVQSYNGDSYMWGNLQAGQY-----VVLNVPNRLVYSAHDYAT-SV 285
QY 273 YPNYWGONAD-----WCTQW---LKHIANAAAGKPTILEEFGWQTPDRDSVYQW 321
DB 286 YPTWFSDPTFPNMPFCIMNKNNGYLFNQNIA-----PVLGEGF--FTLOSTTDQW 336
QY 322 TOT-----VRTNGEAG-----WNFW 336
DB 337 LKTLVOYLRTAQYGADSTQWTFW 360

RESULT 2
GUND_PAPEO
ID GUND_PAPEO STANDARD; PRT; 397 AA.
AC P23548;
DT 01-NOV-1991 (Rel. 20, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DE 15-JUL-1998 (Rel. 36, Last annotation update)
DE Endoglucanase (EC 3.2.1.4) (Endo-1,4-beta-glucanase) (Cellulase).
OS Paenibacillus polymyxa (Bacillus polymyxa).
OC Bacteria; Firmicutes; Bacillales; Paenibacillaceae; Paenibacillus.
OX NCBI_TaxID=1406;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90170877; PubMed=2307659;
RA Baird S.D., Johnson D.A., Seligy V.L.;
RT "Molecular cloning, expression, and characterization of
RT endo-beta-1,4-glucanase genes from Bacillus polymyxa and Bacillus
RT Circulans.";
RL J. Bacteriol. 172:1576-1586(1990).
CC -!- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic
CC linkages in cellulose.
CC -!- SIMILARITY: BELONGS TO CELLULASE FAMILY A (FAMILY 5 OF GLYCOSYL
CC HYDROLASES).
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; M33791; AAA22631.1; -
CC DR PIR; A35136; A35136.
CC DR HSRP; P54583; IECE.
CC InterPro; IPR001547; GH_5.
```

```
DR Pfam: PF00150; cellulase; 1.
DR PROSITE: PS00659; GLYCOSYL_HYDROL_F5; 1.
KW Cellulose degradation; Hydrolase; Glycosidase.
FT ACT_SITE 194 194 PROTON DONOR (HY SIMILARITY).
FT ACT_SITE 317 317 NUCLEOPHILE (HY SIMILARITY).
SQ SEQUENCE 397 AA; 44357 MW; B9C3E802C04F0A2A CRC64;

Query Match
Best Local Similarity 6.7%; Score 138.5; DB 1; Length 397;
Matches 83; Conservative 48; Mismatches 143; Indels 107; Gaps 19;

QY 6 VTASGGFVNLGLPYRYGGTNNYLSYQSHADVDVLLAKAQAAMNLSVIRKTFIDIGSLD 65
DB 47 VDESGKEAFAENGLNMFGLTEPNYTLHGLMSRSMDDMLDOYKKEGYNLRLPYSNULFSS 106
QY 66 GSPTTIDGNKNGFYQYWDPTGAPAYNDGPTGLQGLDYAIAASAAHGLRVIVLVNTWK 125
DB 107 SRPDSIDYHKN-----PDL-----VGLNPITQIMDKLIEKAGQGIQIIL----DRH 148
QY 126 EFGMDQYDKWKYGLPYHDNFYDPTQOAYKNVNH---LLNRVNSITGVTYKNDPTIFA 182
DB 149 RPSGGGSELMYTSQYPS-----RWISDKMKLADR-----YKNNPVTIVG 188
QY 183 WELANERPCVSGTLPSTGCTQATIVNW---VDMSAYVKSIDPNHMSVSGDEGFYIGS 239
DB 189 ADLNEPH--QASWGTGNAST-----DWRLAORAGNALILSVNPNMLILVEGYDINVG 241
QY 240 TQSGMPYNDPSQGVND-NALLRVKNIDFGTYHLY-----PNYWCQ 279
DB 242 NNSQYW-WGNCNLTCVANYPVVLDVPRVYSPHIDYGPVSSQPMFNDPAPSNLPAIMQ 300
QY 280 NADWGTOVI-KHITANAAAGKPTILEEFG-----WQTP-----DRDSVYQW 321
DB 301 T--WG--YISKQNI-----PVLGEGFGRNVLDLSCPECKWONALVHYIGANNLYFTY 349
QY 322 TOTVRTNGEAG-----WNFW 336
DB 350 WSLNPNSGDGTGLLDDDTTW 370

RESULT 3
GUND_CELFI
ID GUND_CELFI STANDARD; PRT; 747 AA.
AC P50400;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DE 15-JUN-2002 (Rel. 41, Last annotation update)
DE Endoglucanase D precursor (EC 3.2.1.4) (Endo-1,4-beta-glucanase)
DE (Cellulase).
DE GN CEND.
OS Cellulomonas fimi.
OC Bacteria; Actinobacteria; Actinobacteriia (class); Actinobacteridac;
OC Actinomycetales; Micrococcales; Cellulomonadaceae; Cellulomonas.
OX NCBI_TaxID=1708;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93209933; PubMed=8458833;
RA Weinke A., Gilkes N.R., Kilburn D.G., Miller R.C. Jr., Warren K.A.;
RT "Cellulose-binding polypeptides from Cellulomonas fimi: endoglucanase
RL J. Bacteriol. 175:1910-1918(1993).
CC -!- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic
CC linkages in cellulose.
CC -!- PATHWAY: Cellulose degradation.
CC -!- SIMILARITY: CONTAINS 2 FIBRONECTIN TYPE III-LIKE DOMAINS.
CC -!- SIMILARITY: CONTAINS 1 BACTERIAL-TYPE CELLULOSE-BINDING DOMAIN
CC (CBD).
CC -!- SIMILARITY: BELONGS TO CELLULASE FAMILY A (FAMILY 5 OF GLYCOSYL
CC HYDROLASES).
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
```


use by non-profit institutions as long as its content is in no way modified and this statement is not removed.

All CC entities requires a license agreement (See http://www.isib-sib.ch/announce/ or send an email to licensee@isib-sib.ch).

CC

DR EMBL; L02544; AAA23089.1; -
HSSP; P07986; IEXG;
InterPro; IPRO01919; Bac_celose-bind.
InterPro; IPRO03961; FN_III.
InterPro; IPRO03962; FNIII_repeat.
InterPro; IPRO01547; GH_5.
Pfam; PF00041; fn3; 2.
Pfam; PF00150; cellulase; 1.
Pfam; PF00553; CBM-2; 1
PRINTS; PR00014; FNTYPEII.
SMART; SMO0060; FN3; 2.
PROSITE; PS00659; GLYCOSYL_HYDROL_F5; 1.
Cellulose degradation; Hydrolase; Glycosidase; Repeat; Signal.
KW SIGNAL 1 39 POTENTIAL.
FT CHAIN 40 747 ENDOGLUCANASE D.
FT DOMAIN 448 542 FIBRONECTIN TYPE-III 1.
FT DOMAIN 546 639 FIBRONECTIN TYPE-III 2.
FT ACT_SITE 208 208 PROTON DONOR (BY SIMILARITY).
FT ACT_SITE 349 349 NUCLEOPHILE (BY SIMILARITY).
SQ SEQUENCE 747 AA; 78936 MW; BD15473C9DBB42BD CRC64;

Query Match 6.2%; Score 128; DH ; Length 747;
Best Local Similarity 21.8%; Pred. No. 0.042;
Matches 63; Conservative 41; Mismatches 119; Indels 66; Gaps 15;

Qy 94 DGPTGLQLDYATASAAAHGLRVIVLTNDWKKEFGMDQDKWYGLPYHDNFVDPRTQQ 153
Db 132 EGKNSLIQFEWLTLCKQGKGVFLDVHSAEADNSG-HVYNMMW-----KCDITTED 182
Qy 154 AYKNWNHLNRVNSITGVTYTKNDDTPIFANELANEPRCVSGTG--LPTSCTQTQAIVNW 211
Db 183 VYEGW-----EWAATRKKODDTIIVGADIINKPHGTGSTERAKWDGTTDKDNPKHF 233
Qy 212 VDQMSAVVKSIDPNHMVSVDGEGRFYIGSTCGSPWPYPDPS--GV---DNNALLRYKNID 266
Db 234 AETASKKLAIAPNLWFV--EGVEIPKGVPWTSTGLTDYYGYTWGGNLRGVRDHPID 291
Qy 267 FGTH-----LYPNWGQ-----NAD-WGTOMIKDHIAAAGKPTEL 304
Db 292 LGARHQDLVYSPhdYGPLVFDFQRKFQFDKASLTADVGNPLFIHDEIDA----PLLI 347
Qy 305 EEFG---WQPTRDSVVYTQTVR---TNCEAGWNFMVL---AGNVNG 344
Db 348 GEWCRLGDGPPODK-----HWALRDLVAEERRLSOTFWLNPNISGDTGC 392

RESULT 4
GUNASTRLI ID GUNA_STRLI STANDARD; PRT; 459 AA.
AC P27035;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-DEC-1992 (Rel. 24, Last sequence update)
DT 01-JUN-1994 (Rel. 29, Last annotation update)
DE Endoglucanase CELA precursor (EC 3.2.1.4) (Endo-1,4-beta-galactosylglucose hydrolase)
GN CELA.
OS Streptomyces lividans.
OC Bacteria; Actinobacteriales; Actinobacteria (class); Actinobacteridae;
OC Actinomycetales; Streptomycinae; Streptomycetaceae; Streptomycetes.
OX NCBI_TaxId=1916;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 28-43.
RC STRAIN=66 / 1326;
RX MEDLINE=92246492; PubMed=1575483;
RA Theberge M., Lacaze P., Shareck F., Morosoli R., Kluepfel D.;
RT "Purification and characterization of an endoglucanase from
Streptomyces lividans 66 and DNA sequence of the gene."
RL Appl. Environ. Microbiol. 58:815-820(1992).

RESULT 5
GUNA_XANCP
ID GUNA_XANCP STANDARD: PRT: 484 AA.
AC P19487:
DT 01-FEB-1991 (Rel. 17, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Major extracellular endoglucanase precursor (EC 3.2.1.4) (Endo-1,4-
DE beta-glucanase) (Cellulase).
GN ENGXA OR XCC3521
OS Xanthomonas campestris (pv. campestris).
OC Bacteria; Proteobacteria; gamma subdivision; Xanthomonas group;
OC Xanthomonas.
OX NCBI_TaxID=340:
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 26-61.
RX MEDLINE=90323605; PubMed=2373365;
RA Cough C.L., Dow J.M., Keen J., Henrissat B., Daniels M.J.:
KT "Nucleotide sequence of the engXA gene encoding the major
RL endoglucanase of Xanthomonas campestris pv. campestris";
RL Gene 89:53-59(1990).
RN [2]
RP SEQUENCE FROM N.A.
KC STRAIN=ATCC 33913 / NCPPB 528;
RX MEDLINE=22022145; PubMed=12024217;
RA da Silva A.C.R., Ferro J.A., Reinach F.C., Farah C.S., Furlan L.R.,
RA Quaggio R.B., Monteiro-Vitorello C.B., Van Sluys M.A., Almeida N.F.,
RA Alves L.M.C., do Amaral A.M., Bertolini M.C., Camargo L.E.A.,
RA Camarotte G., Cannavan F., Cardozo J., Chambergo F., Clapina L.P.,
RA Cicarelli R.M.B., Coutinho L.L., Cursino-Santos J.R., El-Dorfy H.,
RA Faria J.B., Ferreira A.J.S., Ferreira R.C.C., Ferro M.T.T.,
RA Formighieri E.F., Franco M.C., Greggio C.C., Gruber A.,
RA Katsighiana A.M., Kishi L.T., Leite R.P., Lemos E.G.M., Lemos M.V.F.,
RA Locati E.C., Machado M.A., Madalera A.M.B.N., Martinez-Rossi N.M.,
RA Martins E.C., Meidanis J., Menck C.F.M., Miyaki C.Y., Moon D.H.,
RA Moreira L.M., Novo M.T.M., Okura V.K., Oliveira M.C., Oliveira V.R.,
RA Pereira H.A., Rossi A., Sena J.A.D., Silva C., de Souza R.F.,
RA Spinola L.A.F., Takita M.A., Tamura R.E., Teixeira E.C., Tezza R.I.D.,
RA Trindade dos Santos M., Truffi D., Tsai S.M., White F.F.,
RA Setubal J.C., Kitajima J.P.;
RT "Comparison of the genomes of two Xanthomonas pathogens with differing
RT host specificities."
KL Nature 417:459-463(2002).
CC -!- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic
CC linkages in cellulose.
CC -!- MISCELLANEOUS: THE C-TERMINAL REGION OF THE PROTEIN IS NOT CRUCIAL
CC FOR ACTIVITY.
CC -!- MISCELLANEOUS: THE THR/PRO-RICH REGION (ALSO TERMED "HINGE") MAY
CC BE A POTENTIAL SITE FOR PROTEOLYSIS.
CC -!- SIMILARITY: BELONGS TO CELLULASE FAMILY A (FAMILY 5 OF GLYCOSYL
CC HYDROLASES).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: M32700; AAA27612.1;
CC DR EMBL: AE012472; BAM42791.1;
CC DR PIR: JH0158;
CC DR HSP: P54583; IECE.
CC DR InterPro: IPR001547; GIL5.
CC DR Pfam: PF00150; cellulase.1.
CC DR PROSITE: P500659; GLYCOSYL-HYDROL-F5; 1.
KW Cellulose degradation; Hydrolase; Glycosidase; Repeat; Signal.
FT SIGNAL 1 25
FT CHAIN 26 484
FT DOMAIN 375 399
FT ACT_SITE 182 182
FT ACT_SITE 303 303
FT ACT_SITE 303 303

FT CONFLICT 410 484
FT VDSNMGYCNRVQVNTGTASCTWSIAVIVTGTNNAMNA
FT TWSQSGTLRASCVDFNRTLAACATAEFGCAAS -> ASP
FT VVGSARKKLPAASRLACHWVPASSTGHRVVIAPSPVTKRP
FT HAAATERRMRVTRLRRLRNLKGPLPAPAIITTOHAPRL
FT (IN REF. 1).
SO SEQUENCE 484 AA: 52241 MW: 6671AE5BF1B7602A CRC64:
Query Match 6.1%; Score 126; DB 1; Length 484;
Best Local Similarity 21.9%; Pred. No. 0.035;
Matches 92; Conservative 52; Mismatches 133; Indels 144; Gaps 26;
QY 4 GFVTASGGQVFLNGLPYRYGCTNNYLSYQSHADVDVLAKAQAANLSVIRTWGFIIDIGS 63
DB 51 GFET---GNHVMHGLMAR-----NWKDMIVQMGLGFNAVRL-PFCPATL 91
QY 64 LDGSVP-TIDGNKNGFYQWDPSTGAPYNDGPTGLOGLDYATASAAAGLRLVIVLTN 122
DB 92 RSDTMPASIDYSRN-----ADLQGLTSLQILDVIAEFNAKGM--YVLLDH 135
QY 123 DWKEFGGMDQYDKWYGLPYHDNFYTDPRTOQAYKNWVHLLNRVNSITGVTYKNDPTIFA 182
DB 136 HTPDCAGISEL--WY-----TGSYTEAQ-----WLADLFVAN-----RYKNVPYVLG 176
QY 183 WELANPRCVSGTLPSTGCTQATIVNW---VDMSAYVKSIDPNHMSVYDGEFY--- 236
DB 177 LDKLNEPH--GAATWGTGNAAT-----DNKAAERGSAALAVAPKWLJAV--EGITDNP 227
QY 237 IGSTQGS-GW-----PYNDPSGCVNNALLRVKNIDFGTYHILYPNWGON----- 280
DB 228 VCSNNGIFGCGNLOPLACTPLNIPA-----NRLLAPHV-YGPDVEVQSYFNSNFPNN 281
QY 281 --ADWGTQMTKDHIAANAALCKPTILKEFGWQTPDRDSVYOTWTQTV----- 325
DB 282 MPALW-----ERHGFQFAGT-HALLLGEFGKYGEGDARDKTQDALVKYLRKSGINUGF 335
QY 326 -----RTNGEAG-----W-----NFMLAGNVNGQPYPNYDGFNVVYPSSTA 362
DB 336 YWSWNPNSGDTGGILRDDMTSVRODKMTLLRLTLMGTAGTNTPTPTPT-----PIHPY 387
QY 363 T 363
DB 388 T 388
RESULT 6
GUNW_ERMCA STANDARD: PRT: 504 AA.
AC Q59395;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE Endoglucanase V1 precursor (EC 3.2.1.4) (Endo-1,4-beta-glucanase; V1)
DE (Cellulase V1).
GN CELV1.
OS Erwinia carotovora.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Pectobacterium.
OX NCBI_TaxID=554;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SCC3193;
RX MEDLINE=95231512; PubMed=7715600;
RA Mae A., Heikinhelmo R., Palva E.T.;
RT "Structure and regulation of the Erwinia carotovora subspecies
RT carotovora SCC3193 cellulase gene celV1 and the role of cellulase in
RL phytopathogenicity";
RL Mol. Gen. Genet. 247:17-26(1995).
CC -!- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic
CC linkages in cellulose.
CC -!- SUBCELLULAR LOCATION: Secreted.
CC -!- SIMILARITY: BELONGS TO CELLULASE FAMILY A (FAMILY 5 OF GLYCOSYL
CC HYDROLASES).
CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: X79241; CAA55823.1; -
 CC HSSP: O85465; IAH3.
 CC InterPro: IPR001956; CHD_3.
 CC InterPro: IPR001547; GH_5.
 CC Pfam: PF00150; cellulase; 1.
 CC Pfam: PF00942; CBM_3; 1.
 CC ProDom: PD001947; CHD_3; 1.
 CC ProSITE: PS00659; GLYCOSYL_HYDROL_F5; 1.
 CC Cellulose degradation: Hydrolase; Glycosidase; Signal.
 KW SIGNAL 1 31
 FT CHAIN 32 504
 FT DOMAIN 32 334
 FT DOMAIN 335 352
 FT DOMAIN 353 504
 FT ACT_SITE 168 168
 FT ACT_SITE 256 256
 SQ SEQUENCE 504 AA: 54963 MW: 974781565FA CRC64;

 CC Query Match 6.0%; Score 125.5; DB 1; Length 504;
 CC Best Local Similarity 24.6%; Pred. No. 0.04;
 CC Matches 62; Conservative 23; Mismatches 76; Indels 91; Gaps 16;
 CC
 CC 107 ASAAHGLRVIVLTNDWKEGMDQYDKWYGLPYHDFNYTDPRTQAYKRWVHLLNRV 166
 CC 114 AVAAQGLGVVILI--DWHTLS-----DN---DPNT---YKAQAKIFFAEM 151
 CC 167 NSITGVYKNDPTIFANELANEPKVCSTGLPTSGTCTQATIVNWDQMSAY-----V 219
 CC 152 AGL-----YGNSPNVI-YEIANEPN--GS-----VTMNGQIRPYALEVTDTI 190
 CC 220 KSIDPNHMSVGDGEGFYIGSTQSGWPNYNDPSDGVN-----NALLRVKNIDFGCTYHLP 274
 CC 191 RSKDPDLNLIIV-----CGSTWSQDIHDAADNQLPDPNTLYAL-----HFYA 231
 CC 275 NYWGONADMTQWIKDHIANAAIGKPTILEEFQWOTPDGRDS-----VYQTWTQTVRT 327
 CC 232 GTIHC-----QFLRDIRYQAQRGAALFVSE--WGTSDASGNGCPFLPESQTWIDFLNN 282
 CC 328 NGEAGNFWMLA 339
 CC 283 RG-LSWVNWLSLS 293
 CC
 CC RESULT 7
 CC GUN5_THEFU STANDARD; PRT; 466 AA.
 CC AC Q01786;
 CC DT 01-JUL-1993 (Rel. 26, Created)
 CC DT 15-JUL-1999 (Rel. 38, Last sequence update)
 CC DT 15-JUN-2002 (Rel. 41, Last annotation update)
 CC DE Endoglucanase E-5 precursor (EC 3.2.1.4) (Endo-1,4-beta-glucanase E-4)
 CC DE (Cellulase E-5) (Cellulase E5).
 CC GN CELE.
 CC OS Thermomonospora fusca.
 CC OC Bacteria; Actinobacteria; Actinobacteria (class); Actinobacteridae;
 CC OC Actinomycetales; Streptosporangineae; Nocardiopsaceae; Thermobifida.
 CC OX NCBI_TaxID=2021;
 CC RN [1]
 CC RP SEQUENCE FROM N.A.
 CC RC STRAIN=YX;
 CC RX MEDLINE=91258320; PubMed=1904434;
 CC RA Lao G., Chantas G.S., Jung E.D., Wilson D.B.;
 CC RT "DNA sequences of three beta-1,4-endoglucanase genes from
 CC RT Thermomonospora fusca";
 CC RL J. Bacteriol. 173:3397-3407(1991).

RN [2]
 RA REVISIONS.
 RL Lao G., Chantas G.S., Jung E.D., Wilson D.B.;
 RN Submitted (APR-1998) to the EMBL/GenBank/DBJ databases.
 [3]
 RN SEQUENCE OF 137-142 AND 157-166.
 RC STRAIN=YX;
 RA Irwin D.C., Spezio M., Walker L.P., Wilson D.B.;
 RT "Activity studies of eight purified cellulases: specificity,
 RT synergism, and binding domain effects";
 RL Biotechnol. Bioeng. 42:1002-1013(1993).
 CC -1- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic
 CC linkages in cellulose.
 CC -1- PATHWAY: Cellulose degradation.
 CC -1- SIMILARITY: BELONGS TO CELLULASE FAMILY A (FAMILY 5 OF GLYCOSYL
 CC HYDROLASES).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: L01577; AAC09379.1; -
 CC PIR: C42360; C42360.
 CC HSSP: P07986; IEXG.
 CC InterPro: IPR001919; Bac_celose-bind.
 CC InterPro: IPR001547; GH_5.
 CC Pfam: PF00150; cellulase; 1.
 CC Pfam: PF00553; CBM_2; 1.
 CC ProSITE: PS00659; GLYCOSYL_HYDROL_F5; 1.
 CC Cellulose degradation: Hydrolase; Glycosidase; Signal.
 KW SIGNAL 1 36
 FT CHAIN 37 466
 FT ACT_SITE 299 299
 FT ACT_SITE 391 391
 FT ACT_SITE 466 466
 SQ SEQUENCE 466 AA: 49800 MW: 106040822DEF82E CRC64;

 CC Query Match 5.9%; Score 123.5; DB 1; Length 466;
 CC Best Local Similarity 22.1%; Pred. No. 0.052;
 CC Matches 81; Conservative 46; Mismatches 117; Indels 123; Gaps 21;
 CC
 CC 16 NGLPYRGCTNNYLSYQSHADVD---DVLA---KAOAMNLSVIRTWGFIDIGSLDGSVP 69
 CC 182 HGNPVLQRMSTHGIQWFDHCLTDSLDALAYDKADTIRLSM----- 224
 CC 70 TIDGNKGFYFOYWDPTSGAPAYNDGPTGL--OGLDYATASAAHGLRVIVLTNDWKEFG 128
 CC 225 -----YIQ-----EDGYETNPRGFTDRHHQLIDMATARGLYVIV---DWHL-- 262
 CC 129 GMDQYDKWYGLPYHDFNYTDPRTQAYKRWVHLLNRVNSITGYTKNDPTIFAMELANE 188
 CC 263 -LTPGDPIYNLDRAKTFEAEIAQRHASK-----TNLY-----EIANE 299
 CC 189 PRCVSGSLTPTSGTCTQATIVNWDQMSAYKSIDPNHMSVGDGEP-YIGSTQSGWPNY 247
 CC 300 PNGV-----SWASTKSYAEVPIVIRQRPDSVIVITGRWSSLGVSQSGC--- 345
 CC 248 NPSDGVNALLRVKNIDFGCTYHLYPNYWGONADMTQWIKDHIANAAIGKPTIL 304
 CC 346 --PAEIAANP--VNASINMY-AHFY-----AASHRDNYLNALRASELFPVVFV 389
 CC 305 EEFQWOTPDGRDSV--YQTWTQTVRTNGE--AGMFWMLAGNVCQYPNY-DGFN---VY 356
 CC 390 TEFCTYTYTGDNDFQMDRYIDLMAERKIGWTKW-----NYSDDFRSGAVF 437
 CC 357 YPSSTAT 363
 CC 438 QPGTCAS 444

```

RESULT 8
MANR_CAPHI STANDARD; PRT: 879 AA.
AC Q95327;
DT 15-JUL-1998 (Rel. 36, Created)
DT 13-JUL-1998 (Rel. 36, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE beta-mannosidase precursor (EC 3.2.1.25) (Mannanase) (Mannase).
GN MANBA.
OS Capra hircus (Goat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC bovidae; Caprinae; Capra.
OX NCBI_TaxID=9925;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Kidney;
RX MEDLINE=97079658; PubMed=8921369;
RA Leiprandt J.R., Kraemer S.A., Halthcock B.E., Chen H., Dyme J.L.,
RA Cavanagh K.T., Frederici K.H., Jones M.Z.;
RT "Caprine beta-mannosidase: sequencing and characterization of the
RT cDNA and identification of the molecular defect of caprine
RT beta-mannosidosis";
RL Genomics 37:51-56(1996).
CC -I- FUNCTION: EXOGLYCOSIDASE THAT CLEAVES THE SINGLE BETA-LINKED
CC MANNOSE RESIDUE FROM THE NON-REDUCING END OF ALL N-LINKED
CC GLYCOPROTEIN OLIGOSACCHARIDES.
CC -I- CATALYTIC ACTIVITY: Hydrolysis of terminal, non-reducing beta-D-
CC mannose residues in beta-D-mannosides.
CC -I- PATHWAY: PENULTIMATE STEP IN N-LINKED OLIGOSACCHARIDE CATABOLISM.
CC -I- SUBCELLULAR LOCATION: Lysosomal.
CC -I- TISSUE SPECIFICITY: FOUND IN SPLEEN AND TO A LESSER EXTENT IN
CC LIVER. NOT DETECTED IN KIDNEY OR BRAIN.
CC -I- DISORDER: DEFECTS IN MANBA CAUSE BETA-MANNOSIDOSIS, A SEVERE
CC RESULTING IN TREMOR, NYSTAGMUS, ATAXIA AND EARLY DEATH. THE
CC PRIMARY STORAGE PRODUCTS ASSOCIATED WITH THE ENZYME DEFICIENCY
CC ARE THE TRISACCHARIDE MAN-BETA-1-4-GLCNAC-BETA-1-4-GLCNAC AND THE
CC DISACCHARIDE MAN-BETA-1-4-GLCNAC.
CC -I- SIMILARITY: BELONGS TO FAMILY 2 OF GLYCOSYL HYDROLASES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U46067; AAC48665.1;
DR InterPro: IPR001649; GH_2.
DR Pfam: PF02836; Glyco_hydro_2_C; 1.
KW Hydrolase; Glycosidase; Glycoprotein; Lysosome; Signal; Polymorphism.
FT SIGNAL 1 17
FT CHAIN 18 879 BETA-MANNOSEDASE.
FT ACT_SITE 457 457 PROTON DONOR (BY SIMILARITY).
FT CARBOHYD 35 35 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 77 77 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 297 297 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 803 803 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT VARIANT 56 56 R -> Q.
FT VARIANT 340 340 T -> S.
SQ SEQUENCE 879 AA; 101385 MW; 2896F03596B480C CRC64;

Query Match 5.6%; Score 115.5; DB 1; Length 879;
Best Local Similarity 20.4%; Pred. No. 0.43;
Matches 62; Conservative 44; Mismatches 85; Indels 113; Gaps 17;

QY 78 FYFQY-----WDSTCAPYNDGPTG--LOGLDYATASAAHGLRVIVLTNDW 124
DB 341 FYFKINGLPIFLKGNWIP--ADSFQDRVTSMDLRLLLQSVVDANMNAIRV-----W 390
QY 125 KEFGGMDYDKWY-----GLPYHDFNY-----TDPRTQAYKNWVHLLNRVNSITG 171

```

```

Db 391 G--GGIYEODEFYELCDLGLIMIWQDFACALYPTDEDFMDSVREEVTHQVRKL----- 443
QY 172 VTYKNDPTIFAWELANEPKVCSCGLPTSGTCTQATIVNWV-----DUMSAYV 219
Db 444 ---KSHPSIITWSGNNEEA-----ALMMGWYDTKPGYLITVTKDYVTLV 486
QY 220 KSIDPNHIVSVGDE--GFYIGS-----TQSGHPYNDPDCGVNDNALLRVKNIDFCTYH 271
Db 487 KNI--RTIVLEGDDQTRPFIISSTPNGAKTTAEGWLSNPYD-----LNYGDVH 512
QY 272 LYPNYWQONADWGTQWIKDHIANAAIGKPTILFEFGWOTPDSDSVYOTMTOTVTKTNGEA 331
Db 533 FY-DYMSDCNNWRT-----FPKARFVSEYGYQS-----WPSFSTLEKVSSE 573
QY 332 GWNF 335
Db 574 DWSY 577

RESULT 9
BIGA_SALTY STANDARD; PRT: 1953 AA.
AC P25927; P25928; O9XCQ3;
DT 01-MAY-1992 (Rel. 22, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Putative surface-exposed virulence protein bigA precursor.
GN BIGA OR STM3478.
OS Salmonella typhimurium.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Salmonella
OX NCBI_TaxID=602;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 14028;
RA Stojiljkovic I., Valentine P., Heffron F.;
RT "Salmonella typhimurium rhs homolog";
RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=LT2 / SGSC1412 / ATCC 700720;
RX MEDLINE=21534948; PubMed=11677609;
RA McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,
RA Courtney L., Porwollik S., Ali J., Dante M., Du P., Hou S., Layman D.,
RA Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,
RA Ryan E., Sun H., Florea L., Miller W., Stoneking T., Nhan M.,
RA Waterston R., Wilson R.K.;
RT "Complete genome sequence of Salmonella enterica serovar Typhimurium
RT LT2.";
RL Nature 413:852-856(2001).
RN [3]
RP SEQUENCE OF 1-765 FROM N.A.
RC STRAIN=LT2;
RX MEDLINE=911100301; PubMed=19871123;
RA Wu J.Y., Siegel L.M., Kredich N.M.;
RT "High-level expression of Escherichia coli NADPH-sulfite reductase:
RT requirement for a cloned cysG plasmid to overcome limiting shikimate
RT cofactor.";
RL J. Bacteriol. 173:325-333(1991).
CC -I- CAUTION: Ref.3 sequence differs from that shown due to frameshifts
CC in positions 414 and 732.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AF133696; AAD39458.1;
DR EMBL: AE008859; AAL22340.1;

```

```
DR EMBL: M64606; AAA27042.1; ALT_FRAME.
DR EMBL: M64606; AAA27043.1; ALT_FRAME.
DR PIR: C39200; C39200.
DR PIR: D39200; D39200.
DR StyGene: SG10437; bigA.
KW Virulence; Repeat; Signal; Complete proteome.
FT SIGNAL 1 27
FT CHAIN 28 1953
FT
FT DOMAIN 101 252
FT REPEAT 101 103
FT REPEAT 104 113
FT REPEAT 114 122
FT REPEAT 123 133
FT REPEAT 134 144
FT REPEAT 145 155
FT REPEAT 156 166
FT REPEAT 167 177
FT REPEAT 178 188
FT REPEAT 189 199
FT REPEAT 200 210
FT REPEAT 211 221
FT REPEAT 222 232
FT REPEAT 233 243
FT REPEAT 244 252
FT CONFLICT 207 207
FT CONFLICT 514 514
FT CONFLICT 1698 1698
FT CONFLICT 1795 1798
FT CONFLICT 1836 1837
FT CONFLICT 1836 1837
SQ SEQUENCE 1953 AA; 200150 MW; 611B3F1C954D91AE CRC64;

Query Match
Best Local Similarity 5.4%; Score 113; DB 1; Length 1953;
Matches 106; Conservative 47; Mismatches 179; Indels 184; Gaps 23;

Qy 9 SGGQFVL--NGLPYRYGCTNNYLSYQSH-----ADVDDVL----- 42
Db 908 AGGDTLLISDQPGSGGIEVYPLKWTHTFYAMMASDYGVVDNEGATIHLOGAGVYGV 967
Qy 43 --AKAAMNLSVIRTWGFDIGSLDGSVPTIDGNKNGFYFYQWDPST----- 87
Db 968 TASRCKALNEGNI-----YLDGLVPTLDDENNIITSTYWPSSLYLTSNGMVAGST 1018
Qy 88 -----GAPAYNDGPTGLGLOL-----DVAIASAAAHG 113
Db 1019 DAGGDATAGTGNITVNNAGFGMALNGCTAINOGVIITLTDGVTGQADELVGMAALNG 1078
Qy 114 LRV!-----VVLTDNMFEGG--MDQYDKWYGLPYHDNPFY-- 146
Db 1079 GVVINDTSGVINIDADYGAQLSDSSSVYIINNGSINLNGSPMDDTDSHMGCTPTDKIWIQ 1138
Qy 147 -----TDPRTOQ-----AYKNVNHLLNRVNSITGVYKNDPTIFANELANEPKCV 192
Db 1139 SLPGSGDSDDRTSOTGPTTAGTILANYGTETLNGDDVNG-----GW-LYNEAGAS 1187
Qy 193 -----GSGTLPTSGTCTQATIVNWDQMSAYVKSIDPNHMSVSGDEGFY----- 236
Db 1188 LTVNGTIVTNGANALANYGLDADAISTHSLFNEADGSITDILLTNGDVTFFNNGDF 1247
Qy 237 IGS7QSGGWPYNDSDGVNALLRVKNIDFGTYHLYPNYWGQADWGTQWKDHIANAA 296
Db 1248 TGS1AGTSYQOEIVNTG--DMTVAEDGKSLVSGSEFY-----NEEDAT---LTNSGSV 1297
Qy 297 AIGKPTILEEFGWTPDRDSVYQWOTQTV-RTNG-----EAGNFWMLAGVNG 344
Db 1298 ECGGNTIIN-----LTRANSLTQVNSGTITATNGYSATITVNGSNDPKIWNATATGVING 1353
Qy 345 ----QPYPNVD-GFNYYVPSSTATVLASEALAI 375
Db 1354 INPDAPLINLGRGYN-FGNQCTINVGDNVAISGG 1388
```

```
RESULT 10
GUNN_ERWCA STANDARD: PKT: 444 AA.
ID Q59394;
AC Q59394;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE Endoglucanase N precursor (EC 3.2.1.4) (Endo-1,4-beta-glucanase N)
DE (Cellulase N).
DE GN CELN.
OS Erwinia carotovora.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Pectobacterium.
OX NCBI_TaxID=554;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATROSEPLICA FCBR C18;
RX MEDLINE=98299944; PubMed=9636315;
RA Olsen O., Thomsen K.K., Weber J., Duns J.O., Svendsen L., Wegener C.,
RA von Wettstein D.;
RT "Transplanting two unique beta-glucanase catalytic activities into
RT one multi-enzyme, which forms glucose.";
RL Biotechnology 14:71-76(1996).
CC -1- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic
CC linkages in cellulose.
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- SIMILARITY: BELONGS TO CELLULASE FAMILY A (FAMILY 5 OF GLYCOSYL
CC HYDROLASES).
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: L39788; AAC37033.1; --
CC HSP; 085465; IAS3H.
CC InterPro: IPR001956; CBD_3.
CC InterPro: IPR001547; GH_5.
CC Pfam: PF00150; cellulase: 1.
CC Pfam: PF00942; CBM_3: 1.
CC ProDom: PD001947; CBD_3: 1.
CC PROSITE: PS00659; GLYCOSYL_HYDROL_F5; 1.
CC Cellulose degradation; Hydrolase; Glycosidase; Signal.
FT SIGNAL 1 31
FT CHAIN 32 444
FT ACT_SITE 168 168 PROTON DONOR (BY SIMILARITY).
FT ACT_SITE 256 256 NUCLEOPHILE (BY SIMILARITY).
SQ SEQUENCE 444 AA; 48300 MW; F47E4179004CBB43 CRC64;

Query Match
Best Local Similarity 5.4%; Score 112.5; DB 1; Length 444;
Matches 61; Conservative 27; Mismatches 89; Indels 95; Gaps 15;

Qy 87 TCAPAYNDGPTGLGLOGLDYAIAASAAHGLRVIVLTDNMFEGGMDQYDKWYGLPYHDNPFY 146
Db 96 TAADGYISNPLSKANKVKEAFAAASGLGVYIII---DWHILS-----DN--- 135
Qy 147 TDPRTOQAY-KNVNHNLLNRVNSITGVYKNDPTIFANELANEPKCVSGTLPTSGTCTQ 205
Db 136 -DPNYYKAQAKTFFFAEMAGLYGSSPNVIY-----ETANE-----PNGG---- 172
Qy 206 ATIVNWDQMSAY-----VKSIDPNHMSVSGDEGFYIGTSGSGWPYNDPSDGVN--- 256
Db 173 ---VTWNGQIRPYALEVTDITRSKDPDNLIIV-----GTGTWSODLHDAADNQL 218
Qy 257 ---NALLRVKNIDFGTYHLYPNYWGQADWGTQWKDHIANAAAIKPKPTILEEFGWQTPD 313
Db 219 PDPNTLYAL-----HFYAGTHG-----QFLRDRIDYAQSRGAIFVSE--WGTSD 261
Qy 314 RDS-----VYQWOTQTVRTNGEAGNFWML 338
```


DR	PROSITE; PS00659; GLYCOSYL_HYDROL_F5; 1.
KW	Cellulose degradation; Hydrolase; Glycosidase; Signal.
FT	SIGNAL 1 28 POTENTIAL.
FT	CHAIN 29 825 ENDOGLUCANASE C.
FT	ACT_SITE 219 219 PROTON DONOR (BY SIMILARITY).
FT	ACT_SITE 335 NUCLEOPHILE (BY SIMILARITY).
SO	SEQUENCE 825 AA; 92015 MW; A1727DA3D7632617 CRC64;
 Query Match 5.3%; Score 110.5; DB 1; Length 825; Best Local Similarity 19.6%; Pred. No. 0.94; Matches 79; Conservative 44; Mismatches 122; Indels 159; Gaps	
Qy	57 GFDTIGSLDGSSVPTIDGNKNGFYQYWDPSTGAPAYNDG-PTGLCGLDYAIAASAAHGRL I15 I : : I : : : : :
Dd	74 GALQLVEVDGQVTLD-----ODGVPIQLRCM-----STHGLQ I06 I : : I : : : : :
Qy	I16 -----VLIVLTNDW-----KEFCGM-----OYDKKWYCL I39 I I : : : : :
Dd	I07 WFEIVNENAFALANDWSNVIRLALYIGENARYNPDLIEKVAGIELAKENDMYVII I66 I : : I : : : : :
Qy	I40 PYHDFNTDPTTOOAYKNWY-----HLLNRVNSITCVTVKNDDPTIFAMEL I85 I : : I : : : : :
Dd	I67 DWHVHACPDNA-DIYOGGVNDEGEYLCAKOFFLIHAEK-----YPNDPHLI-YEL I216 I : : I : : : : :
Qy	I86 ANEPRVCVSG-----TLPTSCTQTATIV-----NKVDQM--S I216 I : : : :
Dd	I217 ANEPSSSSGGPGITNDEDGWAEVRYAOPLVDLRDSGNAEDNIIVGSPNNKSHMDLA I276 I : : I : : : : :
Qy	I217 AYKSIDPNHWVSGDEGFVIGTSGSGWPYNPDSQDVD-NNALLRVK-NIDFGTYHLYP I274 I : : : : : : :
Dd	I277 AADNPIDDH--TWYTLHFYTGTHEGTSYPEGISSEDRSNVMNAKYALDKGK-AIPA I333 I : : I : : : : : :
Qy	I275 NYWG-----ONADMGTOWIKDHTANAAILCKPTLIEFPCWQ I310 I : : : : :
Dd	I334 TENGVSEADGNGPYLNADVWLNFNLNENNISW-TWLSLTNNKETSAGTFPFILNESDAT I392 I : : I : : : : :
Qy	I311 --TPDRDSVYQWT-QTVRTNGEAGNFWMLAGNVAGCPYNYD I351 I : : I : : : : :
Dd	I393 DLDPGED---QVMSMEELSVSGE-----YVKSRIILGEEYQPID I427 I : : I : : : : :
 RESULT 14	
ID	GUNB_CLOTM STANDARD; PRT: 563 AA.
AC	P04956;
DT	I3-AUG-1987 (Rel. 05, Created)
DT	I3-AUG-1987 (Rel. 05, Last sequence update)
DE	01-FEB-1995 (Rel. 31, Last annotation update)
DE	Endoglucanase B precursor (EC 3.2.1.4) (ECB)
DE	(Cellulase B).
GN	CFLB.
OS	Clostridium thermocellum.
OC	Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
CC	Clostridium.
RX	NCBI_TaxID=1515;
RP	[1]
RC	SEQUENCE FROM N.A.
RC	STRAIN=NCIB 10882;
RX	MEDLINE=86148508; PubMed=3453102;
RA	Grepinet O., Beguin P.;
RT	"Sequence of the cellulase gene of Clostridium thermocellum coding
RL	for endoglucanase B,"
RL	Nucleic Acids Res. 14:1791-1799(1986).
CC	-I- FUNCTION: THIS ENZYME CATALYZES THE ENDOHYDROLYSIS OF 1,4-BETA-
CC	GLUCOSIDIC LINKAGES IN CELLULOSE, LICHENIN AND CEREAL BETA-D-
CC	GLUCANS.
CC	-I- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic
CC	linkages in cellulose.
CC	-I- DOMAIN: A 24 RESIDUES DOMAIN IS REPEATED TWICE IN THIS ENZYME AS
CC	WELL AS IN OTHER C.THERMOCELLUM CELLULOSEOMES ENZYMES. THIS DOMAIN
CC	MAY FUNCTION AS THE BINDING LIGAND FOR THE SL COMPONENT.
CC	-I- SIMILARITY: BELONGS TO CELLULASE FAMILY A (FAMILY 5 OF GLYCOSYL
CC	HYDROLASES).


```
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: X03592; CAA27266.1; -.
CC PIR: A23512; C2CLBM.
CC HSP: P54583; ICEE.
CC InterPro: IPR002105; Dockerin_1.
CC InterPro: IPR002048; EF-hand.
CC Pfam: PF001547; GH_5.
CC Pfam: PF00150; cellulase; 1.
CC Pfam: PF00404; Dockerin_1; 2.
CC PROSITE: PS00018; EF_HAND; UNKNOWN_1.
CC PROSITE: PS00448; CLOS_CELLULOSE_RPT; 2.
CC PROSITE: PS00659; GLYCOSYL_HYDROL_F5; 1.
CC Cellulose degradation; Hydrolase; Glycosidase; Repeat; Signal.
CC SIGNAL 1 27
CC CHAIN 28 563
CC ACT_SITE 204 204
CC ACT_SITE 363 363
CC DOMAIN 502 557
CC REPEAT 502 526
CC REPEAT 534 557
CC SEQUENCE 563 AA; 63929 MW; 866FE55704A1DE4B CRC64;
CC -----
CC Query Match 5.3%; Score 110; DB 1; Length 563;
CC Best Local Similarity 21.8%; Pred. No. 0.65;
CC Matches 52; Conservative 35; Mismatches 81; Indels 70; Gaps 14;
CC -----
CC 83 WDPSTGAPAYNGPTGLOG---LDYATASAAAGLRVTIVLTNDKKEFGMDQYDKWYGL 139
CC 114 YPPSTDTSYNNPALAGLSYELFMLENFRVGIKVLIDVHSPETDQNG-HNTPLMWTNT 172
CC 140 PYHDNEYTPRTOQAYKN---WVNHLLNRVNSITGVYKTDPTIFAWELANEPKCVSGT 196
CC 173 TI-----TEEIFKAWVW-----AERYKNDTDITIGFDLKNKPH-TNNGT 211
CC 197 LPTSGTCTCTOATI-----VWV---VDQMSAYVKSIDPNHMSVGDGEFYIGTQGSWP 246
CC 212 MKIK---AQSAIWDSDSNHPNKKVAKETALAILVHPNVLIFVEGVME----- 258
CC 247 YNDPSGDVNNALLRVK---NIDFGTYHLYPVNGQADMGCTQWIKDHIAAAAAIGK 300
CC 259 ---PKDGIWDETDFTSPWTCNNDY-----YGNWNGN-----LRCKVDYPIN---LGR 301
CC -----
CC RESULT 15
CC MANA_STRL1 STANDARD; PRT; 383 AA.
CC AC P51529;
CC DT 01-OCT-1996 (Rel. 34, Created)
CC DT 30-MAY-2000 (Rel. 39, Last sequence update)
CC DT 30-MAY-2000 (Rel. 39, Last annotation update)
CC DE Mannan endo-1,4-beta-mannosidase precursor (EC 3.2.1.78) (beta-
CC mannase) (1,4-beta-D-mannan mannanohydrolase).
CC GN MANA.
CC OS Streptomyces lividans.
CC OC Bacteria; Actinobacteria; Actinobacteria (class); Actinobacteridae;
CC Actinomycetales; Streptomyces; Streptomyces; Streptomyces.
CC OX NCBI_TaxID=1916;
CC RN [1]
CC RP SEQUENCE FROM N.A., AND SEQUENCE OF 36-42.
CC RC STRAIN=66 / 1326;
CC RA MEDLINE=93207541; PubMed=8457214;
CC RA Arcand N., Kluepfel D., Paradis F.W., Morosoli R., Shareck F.;
CC RT "Beta-mannase of Streptomyces lividans 66: cloning and DNA sequence
CC of the manA gene and characterization of the enzyme.";
CC RT Biochem. J. 290:857-863(1993).
```

```
CC -----
CC [2]
CC REVISIONS TO C-TERMINUS.
CC STRAIN=66 / 1326;
CC Shareck F.;
CC Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: OPTIMAL ENZYME ACTIVITY IS OBTAINED AT 58 DEGREES
CC CELSIUS AND A PH OF 6.8.
CC -!- CATALYTIC ACTIVITY: Random hydrolysis of 1,4-beta-D-mannosidic
CC linkages in mannans, galactomannans, glucomannans, and
CC galactoglucomannans.
CC -!- SUBUNIT: MONOMER.
CC -!- SIMILARITY: BELONGS TO CELLULOSE FAMILY A (FAMILY 5 OF GLYCOSYL
CC HYDROLASES).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: M92297; AAA26710.2; -.
CC InterPro: IPR002883; CHD_5.
CC InterPro: IPR001547; GH_5.
CC Pfam: PF00150; cellulase; 1.
CC Pfam: PF02013; CBM_10; 1.
CC PROSITE: PS00659; GLYCOSYL_HYDROL_F5; 1.
CC Hydrolase; Glycosidase; Signal.
CC SIGNAL 1 35
CC CHAIN 36 383
CC DOMAIN 336 340
CC SEQUENCE 383 AA; 39681 MW; 5DB4B407C64E94C3 CRC64;
CC -----
CC Query Match 5.3%; Score 109.5; DB 1; Length 383;
CC Best Local Similarity 20.5%; Pred. No. 0.44;
CC Matches 90; Conservative 54; Mismatches 113; Indels 181; Gaps 27;
CC -----
CC 1 APAGFTVAGSGQFVLNGLPYRGYGTNNYLSQSHADVDVLLAKAAAMNLSVIRTWGTFD 60
CC 4 ARSLTITTAGMAFAVLGLLALAGPS-----AGRAEA-----AAGGIIH 41
CC 61 IGSLOGSVPTIDGNKNGFYFO-----YMDPSTGAPAYNDGPTGIQGLDYAIAASAAHGL 114
CC 42 VS--NGRV--VEGNGSAFMRGVNHAIYTWYDPR-----TG-----SIADIAAKGA 82
CC 115 RVI-VVLTNDKHEFGMDQYDKWYGLPYHDNEYTPRTOQAYKNVHLLNRVNSITGV 173
CC 83 NTVRVVJSS-----GG-----RWTKYSASEVSALICOC 110
CC 174 YKNDPTIFAWELANEPKCV-----GSGTLPTSGTCTOATIVNWVDMSAYVKS 221
CC 111 -----KANKVICVLEVDHTTGYGKDGATSLDQNGD-----YWGVKSAAMRA 152
CC 222 IDPNHMSVGDGEFGYIGTQGSWPYNDPSGVNUNALLRVKNIIDFGTYHLY--PNYWGQ 279
CC 153 QEDYVVVNIQNEPF--GNTNYAAM-----TDATKSAICKRLGAGLCHALMVADPN-WQO 203
CC 280 NADW-GTQ-----WIKDHIAAAAAIGKPTILEEFG 308
CC 204 --DWSGTMRSNAASVFAADPDPRNTVFSIHMYGVVDYTAAEVRDYLNAFVNGNGLPIVVG 261
CC 309 WQ-----TPDRDSVYQTWTQTVRTNCEAGNFWMLAGNVNGOPY-----PN----- 349
CC 262 DOHSDGNPDDEAIDMAT-AQSLCV-CYLQWS-W--SCNGGGVEYLDWVNGFDPNLSLTSNGN 316
CC 350 ---YDGFNVVYPSSTATV 364
CC 317 RILYGSNGIAATSRATV 334
CC -----
CC Search completed: November 13, 2002, 11:55:03
CC Job time : 11.6207 secs
```

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Kun on: November 13, 2002, 10:50:41 : Search time 34.7522 Seconds
(without alignments)
2223.392 Million cell updates/sec

Title: us-09-917-378-3

Perfect score: 2079

Sequence: 1 APAGFVTASGCGFVLNCLPY.....YYPSTATVLASEALAISTG 375

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREML_21:*

- 1: sp_archaea:*
- 2: sp_bacteria:*
- 3: sp_fungi:*
- 4: sp_human:*
- 5: sp_invertebrate:*
- 6: sp_mammal:*
- 7: sp_mhc:*
- 8: sp_organelle:*
- 9: sp_phase:*
- 10: sp_plant:*
- 11: sp_rodent:*
- 12: sp_virus:*
- 13: sp_vertebrate:*
- 14: sp_unclassified:*
- 15: sp_rvirus:*
- 16: sp_bacteriophage:*
- 17: sp_archaeal:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	829.5	39.9	666	Q9RIK7	Q9rik7 thermotoga
2	812	39.1	680	Q9RIK9	Q9rik9 thermotoga
3	811	39.0	669	Q9X0V4	Q9x0v4 thermotoga
4	787.5	37.9	694	Q9Z187	Q9z187 bacillus st
5	706.5	34.0	439	Q9P893	Q9p893 agaricus bi
6	706.5	34.0	439	Q9Z401	Q9z401 agaricus bi
7	630	30.3	377	Q00012	Q00012 aspergillus
8	599	28.8	578	Q96V96	Q96v96 orpinomyces
9	596	28.7	437	Q99036	Q99036 trichoderma
10	543.5	26.1	431	Q9FJ23	Q9fj23 arabidopsis
11	537	25.8	408	Q9SG95	Q9sg95 arabidopsis
12	532.5	25.6	416	Q9FT03	Q9ft03 coffea arab
13	530	25.5	395	Q9RVL3	Q9rvl3 lycopersico
14	528	25.4	399	Q93WT4	Q93wt4 lycopersico
15	517.5	24.9	411	Q9FZ29	Q9fz29 arabidopsis
16	516	24.8	414	Q9SG94	Q9sg94 arabidopsis

17	498	24.0	448	10	Q9LZV3	Q9lzv3 arabidopsis
18	495	23.8	431	10	Q9M0H6	Q9m0h6 arabidopsis
19	493	23.7	427	10	Q9FT02	Q9ft02 coffea arab
20	491.5	23.6	439	10	Q8SAY2	Q8say2 oryza sativ
21	483	23.2	445	10	Q94J47	Q94j47 oryza sativ
22	476	22.9	468	10	Q8SAY1	Q8say1 oryza sativ
23	464	22.3	401	10	Q9FU06	Q9fu06 lycopersico
24	464	22.3	415	10	Q9FZ03	Q9fz03 lycopersico
25	426.5	20.5	369	10	Q8E540	Q8e540 lycopersico
26	411.5	19.8	403	10	Q9SKU9	Q9sku9 arabidopsis
27	407	19.6	311	10	Q9LW44	Q9lw44 arabidopsis
28	399.5	19.2	442	16	Q9AA10	Q9aa10 caulobacter
29	255	12.3	218	10	Q93X40	Q93x40 lactuca sat
30	193	9.3	171	10	Q9LKG5	Q9lkg5 chlorocella v
31	171.5	8.2	589	16	Q87851	Q87851 streptomyces
32	143	6.9	2817	16	Q97K42	Q97k42 clostridium
33	141.5	6.8	425	2	Q9LAJ3	Q9laj3 clostridium
34	140.5	6.8	649	16	Q8YD80	Q8y80 brucella me
35	140	6.7	367	5	Q8WPJ2	Q8wpj2 mytilus edu
36	139.5	6.7	397	2	Q8RP23	Q8rp23 puenibacilli
37	136.5	6.6	722	16	Q97L56	Q97l56 clostridium
38	135	6.5	660	2	Q9L3J2	Q9l3j2 clostridium
39	134	6.4	482	2	Q9FAC7	Q9fac7 rhodococcus
40	134	6.4	1684	2	Q0J658	Q0j658 unidentified
41	132.5	6.4	727	2	Q9AF65	Q9af65 clavibacter
42	129.5	6.2	741	17	Q87KH5	Q8tkh5 methanobac
43	128.5	6.2	585	2	Q8VUT3	Q8vut3 pseudomonas
44	127	6.1	425	2	Q93Q07	Q93q07 thermus cal
45	126.5	6.1	745	2	Q9K5C7	Q9k5c7 clavibacter

ALIGNMENTS

RESULT 1

Q9RIK7	ID	Q9RIK7	PRELIMINARY;	PNT;	666 AA.
AC	Q9RIK7				
DT	01-MAY-2000	(TREMBLrel. 13, Created)			
DT	01-MAY-2000	(TREMBLrel. 13, Last sequence update)			
DT	01-JUN-2002	(TREMBLrel. 21, Last annotation update)			
DE	Beta-mannosidase (Beta-mannanase).				
GN	MANB OR MAN5.				
OS	Thermotoga neapolitana.				
OC	Bacteria; Thermotogae; Thermotogales; Thermotogaceae; Thermotoga.				
OX	NCBI_TaxID=2337;				
RP	[1]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=5068;				
RC	STRATIN=5068;				
RA	Parker K.N., Lam D., Duffaud G., Shead M.A., Mathur E.J., Kelly R.M.;				
RT	"Amino Acid Sequence of beta-mannosidase genes from the				
RT	hyperthermophilic bacteria Thermotoga maritima and Thermotoga				
RT	neapolitana."				
RL	Submitted (AUG-1998) to the EMBL/GenBank/DBJ databases.				
RN	[2]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=DSM 5068;				
RC	MEDLINE=21473892; PubMed=11590605;				
RA	Parker K.N., Chhabra S.R., Lam D., Callen W., Duffaud G.D.,				
RA	Snead M.A., Short J.M., Mathur E.J., Kelly R.M.;				
RT	"Galactomannanases Man2 and Man5 from Thermotoga species: growth				
RT	physiology on galactomannans, gene sequence analysis, and biochemical				
RT	properties of recombinant enzymes."				
RL	Biotechnol. Bioeng. 75:322-333(2001).				
DR	EMBL; Y17981; CAB56856.1; .				
DR	EMBL; AY033477; AAK53459.1; .				
DR	InterPro: IPR001547; GH5.				
DR	Pfam: PF00150; cellulase_1.				
SQ	SEQUENCE 666 AA; 76695 MW; 7EF986115E4C58CD CRC64;				

Query Match 39.9%; Score 829.5; DB 2; Length 666;
Best Local Similarity 43.5%; Pred. No. 5.5e-50;
Matches 173; Conservative 63; Mismatches 119; Indels 43; Gaps 13;

OS Agaricus bisporus (Common mushroom);
 OC Eukaryota; Fungi; Basidiomycota; Hymenomycetes; Homobasidiomycetes;
 OC Agaricales; Agaricaceae; Agaricus.
 OX NCBI_TaxID=5341;
 KN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=D649; TISSUE=MYCELIUM;
 RX MEDLINE=9717791; PubMed=9025297;
 RA Yague E., Mehak-Zunic M., Morgan L., Wood D.A., Thurston C.F.;
 RT "Expression of CEL2 and CEL4, two proteins from Agaricus bisporus with
 similarity to fungal cellobiohydrolase I and beta-mannanase,
 respectively, is regulated by the carbon source.";
 RT Microbiology 143:239-244(1997).
 RL EMBL: Z50095; CAA90423.1; -;
 DR HSP: P00725; IAZ6.
 DR InterPro: IPR001589; Actbind_actnin.
 DR InterPro: IPR000254; CHD_fungal.
 DR InterPro: IPR001547; GH_5.
 DR Pfam: PF00734; CBM_1; 1.
 DR ProDom: PD001821; CHD_fungal; 1.
 DR SMART: SM00236; fCHD; 1.
 DR PROSITE: PS00019; ACTININ_1; UNKNOWN_1.
 DR PROSITE: PS00562; CHD_FUNGAL; 1.
 KW Glycosidase; Hydrolase; Signal.
 FT SIGNAL 1 19
 FT CHAIN 20 439 CEL4B MANNANASE.
 SQ SEQUENCE 439 AA: 46940 MW: 7250A2D4A273D522 CRC64;
 Query Match 34.0%; Score 706.5; DB 3; Length 439;
 Best Local Similarity 40.8%; Pred. No. 1.3e-41;
 Matches 154; Conservative 57; Mismatches 131; Indels 35; Gaps 12;
 QY 4 GFVTASGGQVFLNGLPYRYGCTNNYY--LSYQSHADVDVLAQAKAMNLSVIRTWGFI 61
 DB 86 GFVKASCTRTLNGQRYTVVGGNSYVWGLTGLSTAMNQAFSDIANAGCTVIRTWGNEV 145
 QY 62 GSLDGSVPTIDGNKNGFYQYWDPTGAPAYNDGPTGLOGLDYAIASAAHGLRVIVLT 121
 DB 146 TS-----PNGNYIQSW--SGARPTINTGASCLLNFEDNVIAAKANGIRLIVALT 192
 QY 122 NDMKEFGMDQYDKWY---GLPYHDNFYDPTQQAQKNNVHLLNRVNSITGVTKNDP 178
 DB 193 NNWADYGGHDVYVQNVGNGOP--HDLFYNTAIDKAFKSYGRAFSR-----YANEP 243
 QY 179 TIFANELANEPRCVSGTLPSTCTCTOATIVNWDQMSAYVKSIDPNHNVSVGDEGFYIG 238
 DB 244 TVNAWELANEPCKGS--TGTTSCCTCTTTTNNWAKEMSAFIKIDSNNHLVAIGDEGFY-N 301
 QY 239 STQGSQWPNDSGDGVDNNALLRVKNIIDFGTYHLYPNYWGONAD---WGTQWIKDHANA 295
 DB 302 QPGAPTYQOG--SEGVDFAENLAISSVDFAFHSYPEPWQGGADAKAWGCTQWITDHAAM 360
 QY 296 AATGKPTILEEFGWQTPDRDSVYQVOTQTVRTNCEAGNFWMLAGNVNGQYPNYDGFNV 355
 DB 361 KRVNKPVILEEFGVTTNQPDY--YAEWFNEIESSGLTGDLIWOAGSHLSTGCDTNP--DGAV 418
 QY 356 YPSSSTATVLASALAI 372
 DB 419 YPDGPVYPLVKSHASAM 435
 RESULT 7
 Q00012 PRELIMINARY; PRT: 377 AA.
 AC Q00012;
 DT 01-NOV-1996 (TrEMBLrel. 01, Created)
 DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
 DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
 DE Mannanase precursor.
 GN MAN1.
 OS Aspergillus aculeatus.
 OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;

OC Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.
 OX NCBI_TaxID=5053;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=KSM 510;
 RX MEDLINE=95078777; PubMed=7987261;
 RA Christgau S., Kauppinen S., Vind J., Kofoed L.V., Dalboge H.;
 RT "Expression cloning, purification and characterization of a beta-1,4-
 mannanase from Aspergillus aculeatus.";
 RL Biochem. Mol. Biol. Int. 33:917-925(1994).
 DR EMBL: L35487; AAA67426.1; -;
 DR InterPro: IPR001547; GH_5.
 DR Pfam: PF00150; cellulase; 1.
 KW Signal.
 FT SIGNAL 1 18 POTENTIAL.
 FT CHAIN 19 377 MANNANASE.
 SQ SEQUENCE 377 AA: 41082 MW: 10E6477555BE3CA2 CRC64;
 Query Match 30.3%; Score 630; DB 3; Length 377;
 Best Local Similarity 35.2%; Pred. No. 2.4e-36;
 Matches 132; Conservative 65; Mismatches 142; Indels 36; Gaps 9;
 QY 1 APAGFVTASGGQVFLNGLPYRYGCTNNYYLSY-QSHADVDVLAQAKAMNLSVIRTWGFI 59
 DB 28 ATTAFPSTSGLHFTIDGKTGYFAGTNSYWGFLTNDDVDLVMSQLAASDLKILRVWGFN 87
 QY 60 DIGS--LDGSVPTIDGNKNGFYQYWDPTGAPAYNDGPTGLOGLDYAIASAAHGLRV 117
 DB 88 DVNKTPTDGTV-----WYQJHANGTSTJNTGADGLOHLDYVVTSAEYGVKLI 135
 QY 118 VLTNDMKEFGMDQYDKWYGLPYHDNFYDPTQQAQKNNVHLLNRVNSITGVTKND 177
 DB 136 INFVNEWTQYGGHQAQYVTAAGAAQTDFYNTAIGAAQYKNYIKAVVSR-----YSSS 187
 QY 178 PTIFANELANEPRCVSGTLPSTCTCTOATIVNWDQMSAYVKSIDPNHNVSVGDEGFY 237
 DB 188 AATFANELANEPRCQ-----CDTSVLYNMISDTSKYIKSLDSKHLVITIGDEGFGL 238
 QY 238 GSTQGSQWPNDSGDGVDNNALLRVKNIIDFGTYHLYPNYWGONADQWIKDHANAA 297
 DB 239 DVSDSGSYPT--VGEGLNFTKNLIGISTIDFGTLHLYPDSNGCTSYDNGNGWITAHAAACKA 297
 QY 298 IGRPTILEEFGWQTPDRDSVYQVOTQTV--RTNCEAGNFWMLAGNVNGQYPNYDGFNV 356
 DB 298 VGRKPCLEEYGV-VTSNHCIVESPWQQTAGNATGISGDLYQYGTTFPSWGQSPN-DGNTFY 355
 QY 357 YPSSSTATVLASEALA 371
 DB 356 YNTSDFECLVTDHVA 370
 RESULT 8
 Q06V96 PRELIMINARY; PRT: 578 AA.
 AC Q06V96;
 DT 01-DEC-2001 (TrEMBLrel. 19, Created)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
 DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
 DE Mannanase ManA.
 GN MANA.
 OS Orpinomyces sp. PC-2.
 OC Eukaryota; Fungi; Chytridiomycota; Neocallimasticales;
 OC Neocallimastacaceae; Orpinomyces.
 OX NCBI_TaxID=50059;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=PC-2;
 RX MEDLINE=21405733; PubMed=11514516;
 RA Steenbakkers P.J.M., Li X.L., Ximenes E.A., Arts J.G., Chen H.,
 RA Ljungdahl L.G., Op Den Camp H.J.M.;
 RT "Noncatalytic docking domains of cellulosomes of anaerobic fungi.";
 RL J. Bacteriol. 183:5325-5333(2001).
 DR EMBL: AF177206; AAL01213.1; -;


```
Db 189 YKDDPTIMAWELNPRCPSPD-----PSGRAIOA-----WITEMAAHVKSLSDRNIIILEAGLE 240
QY 234 GFYIGSTOGSCWPNPSPD-GVDNNALLRVKNIDFGTYHLYPNYWQNA-----DMCTQ 286
Db 241 GFY-GOSSPOSKTLNPPGQGTDFIANNRIPGIDFVTVHSYPDWEPDSSESQSOMDFLNK 299
QY 287 WIKDHIANAA-AIGKPTILEEF-----GWQTPDRDSVYOT-----WTQTVRTNGEAGW 333
Db 300 WLDAAHQDAQNVLUHKPILIAEPKSKMKKPCYTPAQRDIVENTVYSKIYGSAAKKGGAAGG 359
QY 334 NFWMLAGNVNGQPNYPN-DGFNYYPSSTATV 364
Db 360 LFWOLL--VNG--IDNFQDGYGIILSSQSSSTV 387

RESULT 11
Q9SG95 PRELIMINARY; PRT; 408 AA.
AC Q9SG95;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DE Putative (1-4)-beta-mannan endohydrolase.
GN Arabidopsis thaliana (Mouse-ear cross).
OS Arabidopsis thaliana (Eudicotyledons; Magnoliophyta; Tracheophyta;
OC Eukaryota; Viridiplantae; Streptophyta; core eudicots; Rosidae;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC Eurosid II; Brassicales; Brassicaceae; Arabidopsi.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. COLUMBIA;
RA Lin X., Kaul S., Town C.D., Benito M.-I., Creasy T.H., Haas B.,
RA Ronning C.M., Koo H., Fujii C.V., Utterback T.R., Barnstead M.E.,
RA Bowman C.L., White O., Nierman W.C., Fraser C.M.;
RT "Arabidopsis thaliana chromosome III BAC T7M13 genomic sequence."
RL Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AC011708; AAF19559.1;
DR InterPro: IPR001547; GH_5.
DR Pfam: PF00150; cellulase; 1.
DR PROSITE: PS00962; RIBOSOMAL_S2_1; UNKNOWN_1.
KW hydrolase.
SQ SEQUENCE 408 AA; 45376 MW; 86C48FA7D311BA15 CRC64;

Query Match 25.8%; Score 537; DB 10; Length 408;
Best Local Similarity 33.3%; Pred. No. 8.7e-30;
Matches 135; Conservative 74; Mismatches 128; Indels 68; Gaps 16;

QY 1 APA-GFTVATGGQFVLNGLPYRYGGTNNYLSYQ-----SHADVDVLAQAQAMNLSVIR 54
Db 24 APSDGEVSRNGVQFILNGKPFYANGFNAYWLAYEATDPTTRFKITNVFQNAISLGLTIAR 83
QY 55 TWGFDIGSLDGSVPTIDGNKNGFYQYWDPTGAPAYNDGPTGLOGLDYAIASAAHGL 114
Db 84 TWGFRD-GAIYRALQATPS-----YDEQT-----FQGLDFVIAEAKRIGI 123
QY 115 RVTIVLTNDWKEFGMDQYDKWY-----LPYHDNFYTDPTQOAYKKNVHLLNRVNSI 169
Db 124 KLILLVNNHDDYGGKKQYVDWARSKEVSSNDDYRNPVIFKDFYKNHVKTVLNRVNTF 193
QY 170 TGVTYKNDPTIFAWELANEPKRCVSGTLPSTGCTCTQATIVNVWDQMSAYVKSIDPNHVS 229
Db 184 TKVAYKDEPAINAWQLMNEPRC-----GVDKSG-----KTLMDMINEMAPFKVSDPNHLS 235
QY 230 VGDEGYIGST---QSGWPYNDPSDGVNDNALLRVKNIDFGTYHLYPNYWQNAWGCT- 285
Db 236 TGHEGYGSSSPERKNSLPVSNANTVGADFIANHNDIDAIDFASMHCGSGLWFLQDQNSR 295
QY 286 -----QWIKDHIANAAI-GKPTILEFGW--QTP-----DRDSVYOT-----WTQTVRT 327
Db 296 LAFIKWLECHIEDAQNILKKPVILAEFGLGSDTPRYTLANRDCGVTTTYYIIYASAQKG 355
```

```
QY 328 NGEAGNFWMLAGNVNGQPNYPNDG-----FNYYPSSTATVLASEA 369
Db 356 GSAAGALFW-----EVISEGSMNFAGPSSIIILSDKSVNTIIEHA 396

RESULT 12
Q9FT03 PRELIMINARY; PRT; 416 AA.
AC Q9FT03;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DE 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE (1-4)-beta-mannan endohydrolase precursor (EC 3.2.1.78).
GN MANB.
OS Coffea arabica (Coffee).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC Asteridae; euasterids I; Gentianales; Rubiaceae; Ixoroideae; Coffeaceae;
OC Coffea.
OX NCBI_TaxID=13443;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=GRAIN;
RX MEDLINE=21362243; PubMed=11469596;
RA Marraccini P., Rogers W.J., Allard C., Andre M.L., Cailliet V.,
RA Lausanne F., Michaux S.;
RT "Molecular and biochemical characterization of endo-b-mannanase from
RT germinating coffee (Coffea arabica) grains."
RL Planta 213:296-308(2001).
DR EMBL; AJ278996; CAC08442.1;
DR InterPro: IPR001547; GH_5.
DR Pfam: PF00150; cellulase; 1.
DR Glycosidase: Hydrolase; Signal.
FT SIGNAL 1 40
FT CHAIN 41 416 ENDO BETA MANNANASE.
SQ SEQUENCE 416 AA; 46824 MW; E484CCCAFC674D66 CRC64;

Query Match 25.6%; Score 532.5; DB 10; Length 416;
Best Local Similarity 32.3%; Pred. No. 1.8e-29;
Matches 134; Conservative 65; Mismatches 127; Indels 89; Gaps 19;

QY 5 FVTASGGQFVLNGLPYRYGGTNNYLSY-----QSHADVDVLAQAQAMNLSVIRTWGFI 59
Db 44 FVKTRGTEFVNGRPLVNGFNAYWLAYMASDPSTRTKVSTTFOASKYCHNAARTWAFS 103
QY 60 DIGSLDGSVPTIDGNKNGFYQYWDPTGAPAYNDGPTGLOGLDYAIASAAHGLRVIV 119
Db 104 DGG-----YRALQSPG--SYNEDM--FKGLDFVYSEAKKGIHLILT 142
QY 120 LTNDWKEFGMDQYDKW-----YCLPYHDNFYTDPTQOAYKKNVHLLNRVNSITGVTY 174
Db 143 LVNMEGTGGKQYQVQARDQGYLNNDDDFDTDP1VRGYFKNHKIVTLTRINSITGLAY 202
QY 175 KNDPTIFAWELANEPKRCVSGTLPSTGCTCTQATIVNVWDQMSAYVKSIDPNHIVSVGDEG 234
Db 203 KDDPTIFAWELMNEPRCSD-----LSCKAIQ-----DWISEMATHVKSIDSDDLIDIGLEG 254
QY 235 FYTGSTQGSQWP-----YNDPSDGVNDNALLRVKNIDFGTYHLYPNYWQNAWGCTQ 286
Db 255 FY-----GESVPQKEYNPGYQVGTDFISNN---RIVQVDFATIHLYPDQWVPNSNDETQ 306
QY 287 -----WIKDHIANAA-AICKPTILEEF-----GWQTPDRDS-----VYQTVR 326
Db 307 AQFVDRWIKHIDDSKYLLEKPLLLTEFGKSSRSPGVQAKRDAYLSHIVTYIYACAA 366
QY 327 TNGEAGN-FWMLAGNVNGQPNYPN-----DGFNYYPS---STATVLASEAIAIST 374
Db 367 GGCVCGNLPMQWMA-----PGMESWGQGYEIVLEENPSTVGVIAQSNRLSS 414

RESULT 13
QBRVL3
```


ID Q8RVL3 PRELIMINARY: PRT: 395 AA.
AC Q8RVL3;
DT 01-JUN-2002 (TrEMBLrel. 21, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Inactive endo-beta-mannanase.
OS Lycopersicon esculentum (Tomato).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC Asteridae; euasterids I; Solanales; Solanaceae; Solanum.
OX NCBI_TaxID=4081;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. WALTER. TISSUE=RIPENING FRUIT PERICARP;
RA Bourgault R., Bewley J.D.;
RT "A cDNA encoding an inactive endo-beta-mannanase expressed in ripening
RT tomato fruit of the cultivar Walter."
RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AY046589; AAK97759.1; -
SQ SEQUENCE 395 AA: 44837 MW: 90A0DCA6ACD45F9B CRC64;
Query Match 25.5%; Score 530; DB 10; Length 395;
Best Local Similarity 34.6%; Pred. No. 2.6e-29;
Matches 140; Conservative 55; Mismatches 132; Indels 78; Gaps 18;
QY 5 FVTASGCGQFVNLGPRYCGTNNYLSY-----QSHADVDVLAQAAMNLSVIRTWGFI 59
DB 32 FVYTNCTHIFALNGKSLYINGFNAYWLMYIAYDPSTRIKVTNTFQQAQSKYKMNVAWTF 91
QY 60 DIGSLDGSVPTIDGNKNGFYQWDPSTGAP-AYNDGPTGLOGLDYALASAAHGLRVIV 118
DB 92 HGS-----RPLQSAAGVNE--QMFQGLDFVISEAKKYGIHLIM 129
QY 119 VLTNDMKEFGMDQYDKW-----YGLPYHDNFYDTPRTOQAYKNVHLLNRVNSITGVT 173
DB 130 SLVNNDAFCGKKQYVEMAVQRCGLTSDDDFTNPMVKGYKNNVKKVLTTRVNTITKVA 189
QY 174 YKNDPTIFAWELANERPCVSGTLPTSGTCTQATIVNWDQMSAYVKSIDPNHMVSGDE 233
DB 190 YKDDPTILSWELINEPRCPD-----LSG-----KTFQNWVLEMAGYLKSIDSNNHLL 241
QY 234 GFYIGTSGGWPNDPS-----DGVNNALLRVKNIDFCYHLYPNW-----GQAD-- 282
DB 242 GFYGNDR-----QYNPSNYIFGTNFISSN---QVQIDFTTHMYPNQWLPGLTQEAQDK 294
QY 283 WGTQWTKHIAANAAATGKPTILEEFGWOT-----PDRDS-----VYQWTQTVRTNGEA 331
DB 295 WASQWIOVHIDSKMLKPLLIAEFCKSTKTCYTVAKRDNYFEKIYGTIFNCAKSGGPC 354
QY 332 GWN-FWMLAGNVGQPPNY-DGFNVYPS--STATVLAASEALAI 372
DB 355 GUGLFW-----QVLQCGMSSFDGQYVQLQESPSRVSILLQSLAL 395
RESULT 14
Q93WT4 PRELIMINARY: PRT: 399 AA.
AC Q93WT4;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Endo-beta-mannanase (EC 3.2.1.78).
OS Lycopersicon esculentum (Tomato).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC Asteridae; euasterids I; Solanales; Solanaceae; Solanum.
OX NCBI_TaxID=4081;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. TRUST. TISSUE=RIPENING FRUIT PERICARP;
RA Bourgault R., Bewley J.D.;
RT "A cDNA encoding an endo-beta-mannanase expressed in ripening tomato
RT fruit of the cultivar Trust;"

RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AY046588; AAK97760.1; -
DR InterPro: IPR001547; GH.5.
DR Pfam: PF00150; cellulase.1.
KW Glycosidase; hydrolase.
SQ SEQUENCE 399 AA: 45338 MW: 805A7D7866E5FF6B CRC64;
Query Match 25.4%; Score 528; DB 10; Length 399;
Best Local Similarity 34.5%; Pred. No. 3.6e-29;
Matches 140; Conservative 54; Mismatches 134; Indels 78; Gaps 18;
QY 5 FVTASGCGQFVNLGPRYCGTNNYLSY-----QSHADVDVLAQAAMNLSVIRTWGFI 59
DB 32 FVYTDGTHFALNGKSLYINGFNAYWLMYIAYDPSTRIKVTNTFQQAQSKYKMNVAWTF 91
QY 60 DIGSLDGSVPTIDGNKNGFYQWDPSTGAP-AYNDGPTGLOGLDYALASAAHGLRVIV 118
DB 92 HGS-----RPLQSAAGVNE--QMFQGLDFVISEAKKYGIHLIM 129
QY 119 VLTNDMKEFGMDQYDKW-----YGLPYHDNFYDTPRTOQAYKNVHLLNRVNSITGVT 173
DB 130 SLVNNDAFCGKKQYVEMAVQRCGLTSDDDFTNPMVKGYKNNVKKVLTTRVNTITKVA 189
QY 174 YKNDPTIFAWELANERPCVSGTLPTSGTCTQATIVNWDQMSAYVKSIDPNHMVSGDE 233
DB 190 YKDDPTILSWELINEPRCPD-----LSG-----KTFQNWVLEMAGYLKSIDSNNHLL 241
QY 234 GFYIGTSGGWPNDPS-----DGVNNALLRVKNIDFCYHLYPNW-----GQAD-- 282
DB 242 GFYGNDR-----QYNPSNYIFGTNFISSN---QVQIDFTTHMYPNQWLPGLTQEAQDK 294
QY 283 WGTQWTKHIAANAAATGKPTILEEFGWOT-----PDRDS-----VYQWTQTVRTNGEA 331
DB 295 WASQWIOVHIDSKMLKPLLIAEFCKSTKTCYTVAKRDNYFEKIYGTIFNCAKSGGPC 354
QY 332 GWN-FWMLAGNVGQPPNY-DGFNVYPS--STATVLAASEALAI 373
DB 355 GUGLFW-----QVLQCGMSSFDGQYVQLQESPSRVSILLQSLRLS 396
RESULT 15
Q9FZ29 PRELIMINARY: PRT: 411 AA.
AC Q9FZ29;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE T6A9.1 protein (1-4).
GN T6A9.1 OR ATIG02310.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsi.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA Federspiel N.A., Palm C.J., Conway A.B., Conn L., Hansen N.F.,
RA Altafi H., Nguyen M., Lam B., Southwick A., Miranda M., Brooks S.,
RA Buehler E., Chao Q., Chin C., Chlou J., Choi E., Gonzalez A.,
RA Howng B., Johnson-Hopson C., Khan S., Kim C., Koo T., Lee J.M.,
RA Lenz C., Liu A., Liu S., Mukharsy N., Pham P., Sakano H., Shinn P.,
RA Toriumi M., Vaysberg M., Yu G., Ecker J., Theologis A., Davis R.W.;
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA Nguyen M., Karlin-Neumann G., Southwick A., Lam B., Miranda M.,
RA Palm C.J., Bowser L., Jones T., Banh J., Carninci P., Chen H.,
RA Cheuk R., Chung M.K., Hayashizaki Y., Ishida J., Kamiya A., Kawai J.,
RA Kim C., Lin J., Liu S.X., Narusaka M., Pham P.K., Sakano H.,
RA Sakurai T., Satou M., Seki M., Shinn P., Yamada K., Shinozaki K.,
RA Ecker J., Theologis A., Davis R.W.;
RL Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL: AC064879; AAG00883.1; -

```
DR EMBL: AY081352; AAL91241.1; -
DR InterPro: IPR001547; GH_5.
DR Pfam: PF00150; cellulase; 1.
KW Hydrolyase.
SQ SEQUENCE 411 AA: 46290 MW: 872ECB54C4D7F218 CRC64:

Query Match      24.9%; Score 517.5; DB 10; Length 411;
Best Local Similarity 32.3%; Pred. NO. 2e-28;
Matches 130; Conservative 69; Mismatches 132; Indels 71; Gaps 14;

Qy 4 GFVTASGQFVLNGLPYRYGGTNNYYL-----SYQSHADVDDVIKAAQAAMNLSVIRTW 56
Db 28 GFVGHNGTQFVLNGEQVYLNGFNAYMMTTAADTASKGRATVTTALRQASAYGMNVARIW 87
Qy 57 GFIDIGSLDGSVPTIDGNKNGFYFOYWDPSITCAP--AYNDGPTGLQGLDYAIAASAAAHLR 115
Db 88 GF-----NEGDYI-----PLQISPGSYSEDV--FKGLDPVYVEAGRFRNIK 125
Qy 116 VIVVLTDNMKEFGMDQYDKWYGLPYHDNFTDPRTOQAYKNVNHLLNRVNSITGVTYK 175
Db 126 LIISLVNPFEDYGGKKYVENAGLDEPDEFYTN SAVKQFYKNHKVTLTRKNTITGRMYK 185
Qy 176 NDPTIFAWELANEPKCVSGTLPITSGCTQOATIV--NWVDQMSAYVKSIDPNHMVSYGDEG 234
Db 186 DDPTIFSWELINEPRCNDSD-----TASNILQDWVKEMASYVKSIDSNHLLLEIGLEG 236
Qy 235 FYIGSTQSGWMPYN-----DPSDGVNALLRVKNIDFGTYHLYPNYW-----GON 280
Db 237 FY-GESIPERTVYNGGRVLTGDTITNN---QIPDIDEATIHIVPDSWLPQSSRTGEQ 292
Qy 281 ADMGTQWIKDHIANA-AAIGKPTILEEF-----GWOTPPDRDSVYQTWTOTVRTNGEAG 332
Db 293 DTFVDKRWICAHIEDCDNITKKPLLITEFGKSKSYPCFSLEKRNKEFKQYVDVIYDSARAG 352
Qy 333 WN-----FWMLACNVNGQPYPNYDGFVNYPPSPSTATVLASEA 369
Db 353 GSCTGGVFWQLTTRTGLGCGTEVFMQAGPNTTAQLIADQS 394
```

Search completed: November 13, 2002, 11:54:09
Job time : 36.7522 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 13, 2002, 11:49:46 ; Search time 11.5841 Seconds
(without alignments)
952.482 Million cell updates/sec

Title: US-09-917-378-3

Perfect score: 2079

Sequence: 1 APAGFVTASGGQFVLNGLPY.....YPPSTATVLAASEALAISTG 375

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents .AA.*
1: /cgn2_6/ptodata/1/iaa/5A-COMB.pap:*
2: /cgn2_6/ptodata/1/iaa/5B-COMB.pap:*
3: /cgn2_6/ptodata/1/iaa/6A-COMB.pap:*
4: /cgn2_6/ptodata/1/iaa/6B-COMB.pap:*
5: /cgn2_6/ptodata/1/iaa/PCTUS-COMB.pap:*
6: /cgn2_6/ptodata/1/iaa/backfiles1.pap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	812	39.1	663	4	US-09-134-078-61 Sequence 61, Appl
2	812	39.1	680	4	US-09-134-078-25 Sequence 25, Appl
3	630	30.3	377	1	US-08-525-697-2 Sequence 2, Appl
4	179.5	8.6	358	1	US-08-604-913B-11 Sequence 11, Appl
5	179.5	8.6	521	1	US-08-276-213-3 Sequence 3, Appl
6	124.5	6.0	711	4	US-08-961-083-82 Sequence 82, Appl
7	123.5	5.9	430	2	US-08-924-440-2 Sequence 2, Appl
8	121.5	5.8	429	1	US-08-745-977-4 Sequence 4, Appl
9	121.5	5.8	429	3	US-09-040-699A-4 Sequence 4, Appl
10	118	5.7	551	2	US-09-033-537A-1 Sequence 1, Appl
11	117.5	5.7	956	4	US-09-134-078-63 Sequence 63, Appl
12	110.5	5.3	501	4	US-09-465-519-4 Sequence 4, Appl
13	109	5.2	296	1	US-08-507-431-6 Sequence 6, Appl
14	109	5.2	296	3	US-09-116-622-6 Sequence 6, Appl
15	109	5.2	296	4	US-09-219-277-6 Sequence 6, Appl
16	109	5.2	296	4	US-09-599-661-6 Sequence 6, Appl
17	109	5.2	467	2	US-08-727-548-2 Sequence 2, Appl
18	109	5.2	467	4	US-08-945-574-1 Sequence 1, Appl
19	108.5	5.2	634	4	US-09-295-744A-2 Sequence 2, Appl
20	108.5	5.2	10182	4	US-09-134-001C-3159 Sequence 3159, Ap
21	106.5	5.1	616	4	US-09-136-574A-47 Sequence 47, Appl
22	106.5	5.1	879	1	US-08-306-546C-2 Sequence 2, Appl
23	106.5	5.1	879	2	US-08-530-524A-2 Sequence 2, Appl
24	106.5	5.1	1426	4	US-09-136-574A-43 Sequence 43, Appl
25	106	5.1	400	2	US-08-713-298B-2 Sequence 2, Appl
26	106	5.1	400	2	US-08-870-180B-2 Sequence 2, Appl
27	106	5.1	400	3	US-08-814-052-4 Sequence 4, Appl

28	106	5.1	400	3	US-08-812-829-4	Sequence 4, Appl
29	106	5.1	400	4	US-09-226-529-2	Sequence 2, Appl
30	105	5.1	462	2	US-08-870-180B-13	Sequence 13, Appl
31	105	5.1	462	4	US-09-226-529-13	Sequence 13, Appl
32	104.5	5.0	501	4	US-09-465-519-2	Sequence 2, Appl
33	104	5.0	485	4	US-09-291-023A-15	Sequence 15, Appl
34	104	5.0	485	4	US-09-291-023A-15	Sequence 15, Appl
35	104	5.0	603	4	US-09-149-727-6	Sequence 6, Appl
36	104	5.0	832	3	US-08-630-820-7	Sequence 7, Appl
37	103.5	5.0	600	6	5268463-2	Patent No. 5268463
38	103.5	5.0	602	2	US-08-882-704A-5	Sequence 5, Appl
39	103.5	5.0	602	4	US-09-151-957-5	Sequence 5, Appl
40	103.5	5.0	602	6	5432081-2	Patent No. 5432081
41	103.5	5.0	1242	4	US-09-488-270A-2	Sequence 2, Appl
42	102.5	4.9	461	4	US-09-134-001C-5311	Sequence 5311, Ap
43	102.5	4.9	485	4	US-09-291-023A-4	Sequence 4, Appl
44	102.5	4.9	485	4	US-09-291-023A-14	Sequence 14, Appl
45	102.5	4.9	613	4	US-09-149-727-5	Sequence 5, Appl

ALIGNMENTS

RESULT 1
US-09-134-078-61
; Sequence 61, Application US/09134078
; Patent No. 6368844
; GENERAL INFORMATION:
; APPLICANT: Bylina, Edward J.
; TITLE OF INVENTION: GLYCOSIDASE ENZYMES
; NUMBER OF SEQUENCES: 72
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Gray Cary Ware & Freidenrich LLP
; STREET: 4365 Executive Drive, Suite 1600
; CITY: San Diego
; STATE: CA
; COUNTRY: USA
; ZIP: 92121
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: Windows95
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/134,078
; FILING DATE: 13-AUG-1998
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/949,026
; FILING DATE: 10-OCT-1997
; APPLICATION NUMBER: 60/056,916
; FILING DATE: 06-DEC-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Haile, Lisa A.
; REGISTRATION NUMBER: 38,347
; REFERENCE/DOCKET NUMBER: 09010/024002
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 858/677-1456
; TELEFAX: 858/677-1465
; INFORMATION FOR SEQ ID NO: 61:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 663 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; FRAGMENT TYPE: internal
US-09-134-078-61

Query Match 39.1% Score 812; DB 4; Length 663;
Best Local Similarity 42.8%; Pred. No. 4.8e-67;
Matches 169; Conservative 61; Mismatches 121; Indels 44; Gaps 13;
Qy 5 FVTASGGQFVLNGLPYRGYGGTNNYLSYQSHADVDVLAKAQMNLVIRTWGTIDIGSL 64

```

Db 17 FVKVNGKFKALNGKEFRFTGNNYMHYKSNMGIDSVLESARDMGKVLRIWGF-----L 71
Qy 65 DGSVPTIDGNKNGFYQYWDPSGTAPAYNDG---PTGLQGLDYAIASAAAHGLRVIVL 120
Db 72 DGSVCRDN-----TYMHPEPGVFGVPEGISNAQSGFERLDYTVAKAKEIGIKLVIVL 125
Qy 121 TNDWKEFGMDQYDKWYGLPYHDNEYTDPRTOQAYKNWYNHLLNRVNSITGVYTKNDPTI 180
Db 126 VNNWDDFGCMNQYVRWFGCTHDDFYRDEKIKEEYKYYVFLVNVHTYTVGPYREEPTI 185
Qy 181 FAWELANPRCVCSTLPTSGTCTQATIVNVVDQMSAYVKSIDPNHMSVSGDEGF---YI 237
Db 186 FAWELANPRC-----ETDKSGN---TLVWYKEMSSYIKSLDPNHLVAVGDEGFFSNYE 237
Qy 238 GSTQSG---WPYNDPSCVDNNALLRVKNIDFGTYHLYPNYWGON---ADWGTQWIKD 290
Db 238 GFKPYGGEAEWAYNGWS-CVDWKKLLSTETVDFGTHLYPSHGWGSPENYAQWGAKEWIED 296
Qy 291 HIANAAAIKPTILEEFG--WQTP-DRDSVYQWTQTVTRTNGEAGNFWMLAGNVMGQP- 346
Db 297 HIKIAKEIGKPVVLEEGYGIPIKSAVNRNTAIYRLWDLVYDLGCGDGMFWMLAGIGESDR 356
Qy 347 -----YPNYDGNVYPPSSTATVLAASE-ALAISTG 375
Db 357 DERGYYPDYGFRIVNDSDSPEAEELIREYAKLFNTG 391

```

RESULT 2

```

US-09-134-078-25
: Sequence 25, Application US/09134078
: Patent No. 6368844
: GENERAL INFORMATION:
: APPLICANT: Bylina, Edward J.
: TITLE OF INVENTION: GLYCOSIDASE ENZYMES
: NUMBER OF SEQUENCES: 72
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Gray Cary Ware & Freidenrich LLP
: STREET: 4365 Executive Drive, Suite 1600
: CITY: San Diego
: STATE: CA
: COUNTRY: USA
: ZIP: 92121
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Diskette
: COMPUTER: IBM Compatible
: OPERATING SYSTEM: Windows95
: SOFTWARE: FastSeq for Windows Version 2.0
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/09/134.078
: FILING DATE: 13-AUG-1998
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: 08/949,026
: FILING DATE: 10-OCT-1997
: APPLICATION NUMBER: 60/056,916
: FILING DATE: 06-DEC-1996
: ATTORNEY/AGENT INFORMATION:
: NAME: Haile, Lisa A.
: REGISTRATION NUMBER: 38,347
: REFERENCE/DOCKET NUMBER: 09010/024002
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 858/677-1456
: TELEFAX: 858/677-1465
: INFORMATION FOR SEQ ID NO: 25:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 680 amino acids
: TYPE: amino acid
: TOPOLOGY: linear
: MOLECULE TYPE: protein
: FRAGMENT TYPE: internal
US-09-134-078-25

```

```

Query Match 39.1%; Score 812; DB 4; Length 680;
Best Local Similarity 42.8%; Pred. No. 5e-67;
Matches 169; Conservative 61; Mismatches 121; Indels 44; Gaps 13;
Qy 5 FVTASGGQVPLNGLPYRYCGTNNYLSYOSHADVDDVLAKAQAAMNLSVIRTWGTFIDIGSL 64
Db 34 FYKVENGFALNGKEFRFTGNNYMHYKSNMGIDSVLESARDMGKVLRIWGF-----L 88
Qy 65 DGSVPTIDGNKNGFYQYWDPSGTAPAYNDG---PTGLQGLDYAIASAAAHGLRVIVL 120
Db 89 DGSVCRDN-----TYMHPEPGVFGVPEGISNAQSGFERLDYTVAKAKEIGIKLVIVL 142
Qy 121 TNDWKEFGMDQYDKWYGLPYHDNEYTDPRTOQAYKNWYNHLLNRVNSITGVYTKNDPTI 180
Db 143 VNNWDDFGCMNQYVRWFGCTHDDFYRDEKIKEEYKYYVFLVNVHTYTVGPYREEPTI 202
Qy 181 FAWELANPRCVCSTLPTSGTCTQATIVNVVDQMSAYVKSIDPNHMSVSGDEGF---YI 237
Db 203 MAWELANPRC-----ETDKSGN---TLVWYKEMSSYIKSLDPNHLVAVGDEGFFSNYE 254
Qy 238 GSTQSG---WPYNDPSCVDNNALLRVKNIDFGTYHLYPNYWGON---ADWGTQWIKD 290
Db 255 GFKPYGGEAEWAYNGWS-CVDWKKLLSTETVDFGTHLYPSHGWGSPENYAQWGAKEWIED 313
Qy 291 HIANAAAIKPTILEEFG--WQTP-DRDSVYQWTQTVTRTNGEAGNFWMLAGNVMGQP- 346
Db 314 HIKIAKEIGKPVVLEEGYGIPIKSAVNRNTAIYRLWDLVYDLGCGDGMFWMLAGIGESDR 373
Qy 347 -----YPNYDGNVYPPSSTATVLAASE-ALAISTG 375
Db 374 DERGYYPDYGFRIVNDSDSPEAEELIREYAKLFNTG 408

```

RESULT 3

```

US-08-525-697-2
: Sequence 2, Application US/08525697
: Patent No. 5795764
: GENERAL INFORMATION:
: APPLICANT: Christgau, Stephan
: APPLICANT: Andersen, Lene N
: APPLICANT: Kauppinen, Sakari
: APPLICANT: Heldt-Hansen, Hans P
: APPLICANT: Dalboege, Henrik
: TITLE OF INVENTION: AN ENZYME EXHIBITING MANNANASE ACTIVITY
: NUMBER OF SEQUENCES: 15
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: No. 5795764o No. 5795764disk of No. 5795764th America, Inc.
: STREET: 405 Lexington Avenue, 64th Floor
: CITY: New York
: STATE: New York
: COUNTRY: United States of America
: ZIP: 10174-6401
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/525,697
: FILING DATE: 21-SEP-1995
: CLASSIFICATION: 435
: ATTORNEY/AGENT INFORMATION:
: NAME: Harrington, James J.
: REGISTRATION NUMBER: 38,711
: REFERENCE/DOCKET NUMBER: 4004.204-US
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 212-867-0123
: TELEFAX: 212-878-9655
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: DK 0486/93
: FILING DATE: 30-APR-1993
: CLASSIFICATION: 435
: INFORMATION FOR SEQ ID NO: 2:

```

SEQUENCE CHARACTERISTICS:
LENGTH: 377 amino acids
TYPE: amino acid
TOPOLOGY: linear
US-08-525-697-2

Query Match 30.3%; Score 630; DB 1; Length 377;
Best Local Similarity 35.2%; Pred. No. 1.7e-50;
Matches 132; Conservative 65; Mismatches 142; Indels 36; Gaps 9;
QY 1 APAGFVTSAGGQFV-LNGLPYRYGGTN-YQSHADVDDVLAKAQAMNLSVIRTWGFI 59
DB 28 ATAPFSTSLIHTIDGKTYFAGTNSYIGFLTNDDVDLWMSQLAASDLKILRVWGFN 87
QY 60 DIGS--LDGSVPTIDGNKNGFYFOYWDPSGAPAYNDGPTGLQGLDYAIASAAHGLRVI 117
DB 88 DVNTKPTDCTV-----WYQLHANGTSTINTGADGLQRLDYVVTSAEKYGVKLI 135
QY 118 VVLTNDWKKEFGMDQYKKYGLPYHDNFYDPTQOAYKNWVHLLNKNVNSITGVTKND 177
DB 136 INFVNEWTDYGMQAYVTAAGAAQTDFTNTAIAQAAKYNIKAVYSR-----YSSS 187
QY 178 PTIFAWELANPRCVSGTLPTSGTCTQATIVNWVDMSAYVKSIDPNHMSVGDGEFYI 237
DB 188 AAIFAWELANPRCG-----CDTSVLYNWIISDTSKYIKSLDSKHLVTTIGDEGFG 238
QY 238 GSTQSGWPNPDSQVDNALLRVKNIDFGTYHLIPNTYWGQADWGTQWIKDHIANAAA 297
DB 239 DVSDSGSYPT-YGEGLNFTKNLGISTIDFGTLHLYPDSWGTSDWGNWITAHAAACKA 297
QY 298 ICKPTILEEFGWQTPDRDSVYQWTQTV-RTNCEAGWFWMLAGNVNGQPYNDGFENVY 356
DB 298 VKRPLLLEBYG-VTSNHCAYESPWOOTAGNATGISDGLYMQYGTTFESWGQSPN-DGNTFY 355
QY 357 YPSSTATVLAEEALA 371
DB 356 YNTSDPTCLVTDHVA 370

RESULT 4
US-08-604-913B-11
Sequence 11, Application US/08604913B
Patent No. 5712142

GENERAL INFORMATION:
APPLICANT: Adney, William S.
APPLICANT: Thomas, Steven R.
APPLICANT: Himmel, Michael E.
APPLICANT: Baker, John O.
APPLICANT: Chou, Yat-Chen
TITLE OF INVENTION: METHOD FOR INCREASING
NUMBER OF SEQUENCES: 12
CORRESPONDENCE ADDRESS:
ADDRESSEE: National Renewable Energy Laboratory
STREET: 1617 Cole Boulevard
CITY: Golden
STATE: CO
COUNTRY: U.S.A.
ZIP: 80401-3393
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBC PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: ASC II (DOS) text
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/604,913B
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/276,213
FILING DATE: 15-070-1994
ATTORNEY/AGENT INFORMATION:
NAME: Edna M. O'Connor

REGISTRATION NUMBER: 29,252
REFERENCE/DOCKET NUMBER: 95-56
TELECOMMUNICATION INFORMATION:
TELEPHONE: 303/384-7573
TELEFAX: 303/384-7499
INFORMATION FOR SEQ ID NO: 11:
SEQUENCE CHARACTERISTICS:
LENGTH: 358 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: Protein
FEATURE:
NAME/KEY: EI-CAT
US-08-604-913B-11

Query Match 8.6%; Score 179.5; DB 1; Length 358;
Best Local Similarity 24.0%; Pred. No. 1.2e-08;
Matches 92; Conservative 50; Mismatches 129; Indels 113; Gaps 24;
QY 1 APAGFVTSAGGQFV-LNGLPYRYGGTN-----NYVLSYQSHADVDDVLAKAQAMNLSV 52
DB 1 AGGYWHITSGREILDANNVPVRIAGINWFGFETCNVYVHGLMSRDYRSMLDQIKSLGYNT 60
QY 53 IRTWGFIDIGSLDGSVPTIDGNKNGFYFOYWDPSGAPAYNDGPTGLQGLDYAIASAAH 112
DB 61 IRL-PYSDDILKPTMP-----NSINFYQMNQD-----LOGLTSLQVMDKIVAYAGO 107
QY 113 GLRIVVLTNDWKKEFGMDQYKKYGLPYHDNFYDPTQOAYKNWVHLLNKNVNSITGV 172
DB 108 GLRII--LORHRPDCSG--QSALWY-----TSSVSEATWISDL-----QALAQ 146
QY 173 TYKNDPTIFAWELANPR-----CVSGTLPTSGTCTQATIVNW-----VDMASAYVKSIDPNH 226
DB 147 RYKGNPTVVGFDLHNEPHDPACWCGCD-PS-----IDWRLAERAGNAVLSVNPNL 196
QY 227 MYSV-----GDEGPIYIGSTGSG-WPYNDPSGDVDNALLRVKN-----IDFGTYHL 272
DB 197 LIFVEGVQSYNGSYWKGNGNLOGAGQVP-----VVLNVPNRLVYSAHDYAT-SV 244
QY 273 YPNWYGNAD-----WCTQW---IKDHIAANAAIGKPTILEEFGWQTPDRDSVYQTW 321
DB 245 YPOTWFSDPTFPNMPGIMNKNMGYLFNQINIA-----PVMJGEFG--TTLOSTTDQTW 295
QY 322 TOT-----VRTNGEAG-----WNFW 336
DB 296 LKTLVOYLRLPTAGYGADSFQWTFW 319

RESULT 5
US-08-276-213-3
Sequence 3, Application US/08276213
Patent No. 5536655
GENERAL INFORMATION:
APPLICANT: Thomas, Steven
APPLICANT: Laymon, Robert
APPLICANT: Himmel, Michael
TITLE OF INVENTION: GENE ENCODING FOR THE EI ENDOGLUCANASE
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESSEE: National Renewable Energy Laboratory
STREET: 1617 Cole Boulevard
CITY: Golden
STATE: CO
COUNTRY: USA
ZIP: 80401-3393
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/276,213

```

; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: O'Connor, Edna
; REGISTRATION NUMBER: 29,252
; REFERENCE/DOCKET NUMBER: NREL IR# 94-08
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (303)231-1000
; TELEFAX: (303)231-1098
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 521 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FRAGMENT TYPE: N-terminal
; US-08-276-213-3

Query Match      8.0%  Score 179.5; DB 1; Length 521;
Best Local Similarity 24.0%; Pred. No. 2e-08;
Matches 92; Conservative 50; Mismatches 129; Indels 113; Gaps 24;

QY 1 APAGFTVSSGQFV-LNGLPYRGCTN-----NYYLSYQSHADVDVLAKAQMNLVS 52
Db 1 AGGGYHTSREILDANNVPRIAGINWFGFETCNVYVHGLWSRDYRSLDIOIKSLGYNT 60

QY 53 IRTWGFIDICSLDGSVPTIDGNKGFYFOYWDPTGAPAYNDGPTGLQGLDYATASAAH 112
Db 61 IRL-PYSDDLKPGTMP-----NSINFQMNQD-----LQGLTSLQVMDKIVAYAGQI 107

QY 113 GLRVIVVLTNDWKEFGGMDQYDKWGLPYLHDNFYTDPTQOAYKNVNHLLNRVNSITGV 172
Db 108 GLRLI--LDRHRPDCS--QSALWY-----TSSVSEATWISDL-----QALAQ 146

QY 173 TYKNDPTIFAWELANERPR---CYSGSLPTSGTCTQATIVNW---VDQMSAYKSIDPNH 226
Db 147 KYGNPTVWGFDLHNEPHDPACWCGGD-PS-----IDWRLAERAGNAVLVSVPNL 196

QY 227 MVSV-----GDEGFYIGSTQSG-WPYNDPDSQVDNNALLRVKN-----IDFGYHL 272
Db 197 LIFVEGVQSYNGDSYWWGNLOGAGYF-----VVLVNPNRLVYSAHDYAT-SV 244

QY 273 YPNWQGNAD-----WGTQM---TKDHIANAAAI GKPTILEEFGWQTPDRDSVYQTM 321
Db 245 YPQTFSDPTFPNNMPCGIWKNKMGYLFNQIA-----PWLCEFG--TTLQSTTDQTM 295

QY 322 TOT-----VRTNEAG-----WNFW 336
Db 296 LKTLVQLRPTAQYGADSFQWTFW 319

RESULT 6
US-08-961-083-82
; Sequence 82, Application US/08961083
; Patent No. 6159469
; GENERAL INFORMATION:
; APPLICANT: Choi et. al.
; TITLE OF INVENTION: Streptococcus pneumoniae Antigens and Vaccines
; NUMBER OF SEQUENCES: 452
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2

```

```

; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/961,083
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Brookes, A. Anders
; REGISTRATION NUMBER: 36,373
; REFERENCE/DOCKET NUMBER: PB340P2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 82:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 711 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-961-083-82

Query Match      6.0%  Score 124.5; DB 4; Length 711;
Best Local Similarity 19.7%; Pred. No. 0.0041;
Matches 72; Conservative 48; Mismatches 126; Indels 119; Gaps 17;

QY 65 DGSVPTIDGNKN--GFYFOYWDPTG-----APAYNDGPTGLQ----- 100
Db 371 DGQL--VDAKKDLFGYRYHWTNPNEGFSLNGERIKFHGCVSLHHDHGALGAENEYKAEYRK 428

QY 101 -----GLDYATASAAHGLRVIVVLTNDWKEFGGMDQYDKWGLPYH 142
Db 429 LKQMKEMCVNSIRTHNPASQTLQIAELCLLVQEEAFDTW--YGGKKPYD--YG-RFF 483

QY 143 DNFTDPTQOAYKNVNHLLNRVNSITGVTYKNDPTIFAWELANERPCVSGTLPTSGT 202
Db 484 EKDATHEARKGER-WSDFDLRTWVE-----RGKNPAIFMMSIGNE---ICEANGDAHSL 535

QY 203 CTQATIVNWDOMSAYKSIDPNHMSVGDGCFYIGSTQSGWPNYNDPSDGV-----DN 256
Db 536 AT-----VKRLVKVTKDVKTRYTMGADKFRFGNGSGGHEKIADELDAVGFNYSEDN 588

QY 257 NALLRVKNIDF-----GTVHL-----YPNY-----WQON 280
Db 589 YKALRAKHPKWLIIYGSETSSATRTGSSYRPERELKHNSGPERNYEQSDYGNDRVGCMT 648

QY 281 ADWGTQWIKDHIAAAA-----ICKPTILEEFGWQTPDRDSVYQTMOTVTRTNGPAG 332
Db 649 ATASWTFDRDNAGYAGOFIWTGTDYIGEPT-----PWHNQNOTPVKSSYFGIVDTAGIPK 703

QY 333 WNFWM 337
Db 704 HDFYL 708

RESULT 7
US-08-924-440-2
; Sequence 2, Application US/08924440
; Patent No. 5871550
; GENERAL INFORMATION:
; APPLICANT: Frits et al.
; TITLE OF INVENTION: MUTANT THERMONOSPORA SPP. CELLULOSE
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genencor International, Inc.
; STREET: 925 Page Mill Road
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:

```

```

; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FASTSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/924,440
; FILING DATE: August 27, 1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Christopher L. Stone
; REGISTRATION NUMBER: 35,696
; REFERENCE/DOCKET NUMBER: GC388
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (650) 846-7555
; TELEFAX: (650) 845-6504
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 430 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
;
US-08-924-440-2

Query Match          5.9%: Score 123.5; DB 2; Length 430;
Best Local Similarity 22.1%: Pred. No. 0.0025;
Matches 81; Conservative 46; Mismatches 117; Indels 123; Gaps 21;

Qy 16 NGLPYRGCTNNYLSYOSHADV---DVLA---KAQAMNLSVIRTWGFIDIGSLDGSVP 69
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 146 HCNVQLRGMSHTGIQWFDHCLTDSLLDALAYDKADIIRLSM----- 188

Qy 70 TIDGNKNGFYFOYWDPTSGAPAYNDGPTGL-QGLDXYATASAAHGLRVIVLTNDWKEFG 128
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 189 -----YIQ-----EDGYETNPRGFTDRHQHQLDMATARGLYVIV---DWHI-- 226

Qy 129 GMDQYDKWYGLPYHDNYFDTPRTOQAYKNMYNHLNVRNSITGVTYKNDPTIFAWELANE 188
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 227 -LTPCDPHYNLDRAKTFEAEIAQRHASK-----TNVLY-----ETANE 263

Qy 189 PRCVGSGTLPTSGTCTQATIVNWVDQMSAYVKSIDPNHMSVGDGEF-YIGSTGSGMPY 247
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 264 PNGV-----SWASIKSYAEVIVPIRQRPDPSVILVTRGWSLLGVSECSG--- 309

Qy 248 NDPDSDGVNALLRVKNIDFCTYHILYPNYWQGNADWGTQWIKKHANA---AATGKPTIL 304
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 310 --PAEIAANP--VNASNIMY-AFHEY-----AASHRDYLNALREASELFPVFV 353

Qy 305 EEFQWQTPDRDSV--YQTWTQTVRTNGE--AGWNFWMLAGNVNGQYPNY-DGFN---VY 356
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 354 TEFCTETVTGDFQANDFQADRYIDLMAERKIGTWTKW-----NYSDDFRSGAVF 401

Qy 357 YPSSTAT 363
   | | | |
Db 402 QPGTCAS 408

RESULT 8
US-08-745-977-4
; Sequence 4, Application US/08745977
; Patent No. 5770406
; GENERAL INFORMATION:
; APPLICANT: Kofod, Lene V.
; APPLICANT: Andersen, Lene N.
; APPLICANT: Kauppinen, Markus S.
; APPLICANT: Christgau, Stephan
; APPLICANT: Dalboge, Henrik
; APPLICANT: Olsen, Hans S.
; TITLE OF INVENTION: An Enzyme With Beta-(1-6)-Endoglucanase
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:

```

```

; ADDRESSSEE: NO. 57704060 No. 5770406disk of No. 5770406th America, Inc.
; STREET: 405 Lexington Avenue, 64th Floor
; CITY: New York
; STATE: New York
; COUNTRY: United States of America
; ZIP: 10174-6401
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/745,977
; FILING DATE: 08-NOV-1996
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
; NAME: Lambiris, Elias J.
; REGISTRATION NUMBER: 33,728
; REFERENCE/DOCKET NUMBER: 4175.204-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-867-0123
; TELEFAX: 212-878-9655
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 429 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
;
US-08-745-977-4

Query Match          5.8%: Score 121.5; DB 1; Length 429;
Best Local Similarity 22.6%: Pred. No. 0.0038;
Matches 85; Conservative 46; Mismatches 142; Indels 103; Gaps 18;

Qy 27 NYILSYOSHADVDDVLAKAQAMNLSVIRTWGFIDIGSLDGSVPTIDGNKNGFYFOYWD-- 84
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 106 NHYRDWINPATVQSV-----HDVGLNTIR---IPIG-----YMSVN 138

Qy 85 --PSTGAPAYNDGPTGLQGLDYATASAAHGLRVIVLTNDWKEFGHDDQYDKWYGL-PY 141
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 139 AIVDTASEPFDAGNLQLPDLDAVVOKAADLGIYVJII---DLHGAPGQGOQDAFTGONPN 194

Qy 142 HDNFYTDPRTOQAYKNMYNHLNVRNSITGVTYKNDPTIFAWELANEP--RCVGSGTLPT 199
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 195 PAGYNTYDYGRAEK-WLSWNTNRIH--TNPAYS---TVGMIEVLNEPVSRIIDGGGRPA 248

Qy 200 SGTCTQATIVNWVDQMSAYVKSIDPNHMSVGDGEF-----YIGSTGSGMP 246
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 249 PG-----QDPSMVQTYTPGALKAVRDAEALNVPSNKKLHVQFNSSKKWDSGDP 296

Qy 247 -----YNDPSDGVNALLRVKNIDFCTYHILYPNYWQGNADWGTQWIKKHIA----- 293
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 297 RSNAAVKNDPVMGFDD-----HNYIGFALQNT--GDQYSLMIHSACTDSRVV 340

Qy 294 ---NAAATGKPTILEEFGWQTPDRDSVYQTVTQTV-RTNGEAGWNFWMLAGNVNGQYP- 348
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 341 SQDFAITGEWSWTSAGADWDHGDGNEFTKFTTAQQQLYESPGHGDGHIYTWKTELNDPRWTY 400

Qy 349 NYDGFNVYPSSTATV 364
   | | | | | | | |
Db 401 SYATYLYNIPTNAAL 416

RESULT 9
US-09-040-699A-4
; Sequence 4, Application US/09040699A
; Patent No. 6022723
; GENERAL INFORMATION:
; APPLICANT: Kofod, Lene V.
; APPLICANT: Andersen, Lene N.
; APPLICANT: Kauppinen, Markus S.
; APPLICANT: Christgau, Stephan
; APPLICANT: Dalboge, Henrik

```

APPLICANT: Olsen, Hans S.
TITLE OF INVENTION: An Enzyme With Beta-(1-6)-Endoglucanase
TITLE OF INVENTION: Activity
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESS: No. 60227230 No. 6022723disk of No. 6022723th America, Inc.
STREET: 405 Lexington Avenue, 64th Floor
CITY: New York
STATE: New York
COUNTRY: United States of America
ZIP: 10174-6401
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/040,699A
FILING DATE: 18-March-1998
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Gregg, Valeta A.
REGISTRATION NUMBER: 35,127
REFERENCE/DOCKET NUMBER: 4175.214-US
TELEPHONE: 212-867-0123
TELEFAX: 212-878-9655
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 429 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-040-699A-4

Query Match 5.8%; Score 121.5; DB 3; Length 429;
Best Local Similarity 22.6%; Pred. No. 0.0038;
Matches 85; Conservative 46; Mismatches 142; Indels 103; Gaps 18;
QY 27 NYLVSQSHADVDVLAQAKAMNLSVIRTWGFDIGSLDGSVPTIDGKNKNGFYQYWD-- 84
Db 106 NHYRDWINPATVQSV-----HDVGLNTR-----IPIG-----YMSYN 138
QY 85 --PSTGAPAYNDGPTGLQLDYAIAASAAHGLRVIVVLTNDWKEFGMDQYDKWYGL-PY 141
Db 139 AIVDTASEPADGNLQPLDVAVQKAADLGIVII-----DLHGAPGGQQQDAFTGQNP 194
QY 142 HDNFYDPTQQAYKNVHLLNRVNSITGVYTKNDPTIFAWELANEP--RCVGSGLPT 199
Db 195 PAGFYNTDYGRAEK-WLSWMTNRH--TNPAYS---TVGMIEVLNPNVSRHDGGRYP 248
QY 200 SGTCTQATIVNWDQMSAYVKSIDPNHMSVGDGEF-----YIGSTQSGWP 246
Db 249 PG-----QDPSPVQTYTPGALKAVRDAEALNVPNNKLLHVQFMSKNWSDGP 296
QY 247 -----YNDPSGDVNNALLRVNIDEGTYHLYPNYQGNADWGTQWIKDHIA----- 293
Db 297 RSNAVKNDPMVGFDD-----HNYIGFALONT--GDOYSLMHSACTDSRVV 340
QY 294 ---NAAIGKPTILEERGMOTPDSDSYVQWTOTV-RTNGEAGNFWMLAGNVNGQYP- 348
Db 341 SGQDFALTGWSMTSGADWHDGNFFTKFTTAQOOLYESPGMDGWIYTWKTELNDPRWTY 400
QY 349 NYDGFNYYPSSTATV 364
Db 401 SYATYLNLYIPTNAAL 416

RESULT 10

US-09-033-537A-1
Sequence 1, Application US/09033537A
Patent No. 5958083
GENERAL INFORMATION:

APPLICANT: Onishi, Masahiro
APPLICANT: Fich, Merete
APPLICANT: Toft, Annette Hanne
APPLICANT: Shlein, Martin
TITLE OF INVENTION: Prevention Of Back-Staining
TITLE OF INVENTION: In Stone Washing
NUMBER OF SEQUENCES: 1
CORRESPONDENCE ADDRESS:
ADDRESS: No. 59580830 No. 5958083disk of No. 5958083th America, Inc.
STREET: 405 Lexington Avenue
CITY: New York
STATE: NY
COUNTRY: U.S.A.
ZIP: 10174
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/033,537A
FILING DATE: 02-MAR-1998
CLASSIFICATION: 008
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 0993/95
FILING DATE: 08-SEP-1995
APPLICATION NUMBER: PCT/DK96/00364
FILING DATE: 03-SEP-1996
ATTORNEY/AGENT INFORMATION:
NAME: Green, Reza
REGISTRATION NUMBER: 38,475
REFERENCE/DOCKET NUMBER: 4492.204-US
TELEPHONE: 212-867-0123
TELEFAX: 212-878-9655
TELEX:
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 551 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-09-033-537A-1

Query Match 5.7%; Score 118; DB 2; Length 551;
Best Local Similarity 18.7%; Pred. No. 0.011;
Matches 81; Conservative 47; Mismatches 119; Indels 186; Gaps 19;

QY 89 APAYNDGPTGLQLDYA-----IASAAHGLRVIVVLTNDWKEFGMDQYDKWY-- 138
Db 1 APAPFGOLKVGQNLVGSQAOVLVGMSSHGL-----QWYGNF 40
QY 139 -----LPYHNFYTDPTQOAYKNV--HLL 163
Db 41 VNKSSLOWMRDNMGINVFRAAWYTAEDGYITDPSVKKNKKEAVQASIDGLYVITDWHIL 100
QY 164 NRVNSIT-----GVYTKNDPTIFAWELANEPRCVSGTCLPTSGTCTQATVNW 211
Db 101 SDGNPTYKAQSKRAFFQEMATLYGNTPNVI-YEIANEP-----NGVSWADVKSY 149
QY 212 VDOMSAYVKSIDPNHMSVGDGE-----GFYIGS----- 239
Db 150 AEEVITAIRAIDPDGVVIVGSPWSDIHLAADNPVSHSNVMYALHEYSCTHGFRLDR1 209
QY 240 -----TQGSWPYNDPS-DCVDNALLRVKNIDFGTYHLYPNY---- 276
Db 210 TYAMNKGAAIEVTEWGTSDASGNGGPFPOSKEWID----FLNARKISWVMNSLADKVETS 266
QY 277 -----W--GONADWGTQWIKDH1-----ANAAIGKPTILEEFQWQTPDR 314
Db 267 AALMPGASPTGGWTDALDSESG-KWVRDQIRQATGGGSGNPTAPAAPTNLS-----ATAGN 321
QY 315 DSIVQTW-----TOTVRTNGEAGNFWMLAGNVNGQYPN----YDGFNYY--PSST 361

Db 322 AOVSLTWNVAVSGATSVTVRRATTSVGGPYTNVATGVGTATSVTNTGLTNGTIIYVWSASNS 381
QY 362 ATVLASEALAIST 374
Db 382 AGSSANSAQASAT 394
RESULT 11
US-09-134-078-63
; Sequence 63, Application US/09134078
; Patent No. 6368844
; GENERAL INFORMATION:
; APPLICANT: Bylina, Edward J.
; TITLE OF INVENTION: GLYCOSIDASE ENZYMES
; NUMBER OF SEQUENCES: 72
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Gray Cary Ware & Freidenrich LLP
; STREET: 4365 Executive Drive, Suite 1600
; CITY: San Diego
; STATE: CA
; COUNTRY: USA
; ZIP: 92121
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: Windows95
; SOFTWARE: FASTSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/134,078
; FILING DATE: 13-AUG-1998
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/949,026
; FILING DATE: 10-OCT-1997
; APPLICATION NUMBER: 60/056,916
; FILING DATE: 06-DEC-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Halle, Lisa A.
; REGISTRATION NUMBER: 38,347
; REFERENCE/DOCKET NUMBER: 09010/024002
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 858/677-1456
; TELEFAX: 858/677-1465
; INFORMATION FOR SEQ ID NO: 63:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 956 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; FRAGMENT TYPE: Internal
US-09-134-078-63
Query Match 5.7%; Score 117.5; DB 4; Length 956;
Best Local Similarity 19.2%; Pred. No. 0.028;
Matches 76; Conservative 52; Mismatches 110; Indels 157; Gaps 19;
QY 3 AGFTVATSGQFVLNGLPY-RYGGTNNYLSYQSHADVDVLAQAQAMNLSVIRTWGFI 61
Db 48 SEGITSLAG-----NSLFSNAGDTSDFY-----NAETVDFL--AENWNSLIR-----IAM 92
QY 62 GSLDGSVPTIDGNKNGFYQYWDPTGAPAYNDGPTGLQG-LDYATASAAAHGLRVI 120
Db 93 GVKEN-----WD---GGNGYIDSPQEQEAKIRKVIDAAIANGIYVII-- 131
QY 121 TNDWKEFGMDQYDKWYGLPYHDNFYDPTQOAYKNWNHLLNRVNSITGVTYK--NDP 178
Db 132 --DWH-----THEAEYTDNAV-----DFFETRMADLYGDTPPNMYEINPEP 170
QY 179 TIFAWELANEPKRCVSGTGLPTSGTCTQATIVNWDQMSAYVKSIDPNHNVSVGDCGFYIG 238
Db 171 IQSNPVP-----IKNYAEQVIAGIRSKOPDNLIIVGTSNYSQO 208

QY 239 STOGSGWPYND-----PSD-----GVDNNALLRVKNIDFGTYHLYPNY 276
Db 209 VDVASADPISDTWVAYTLHFYAAFNPNDNLNRVAQTALDNNVALFV----- 254
QY 277 WQONADWGTOMIKDHIANAAAIKPTILEFEGWOTPDORSVYQVOTWTOTVRTNGEAGNFW 336
Db 255 -----TEWGT-----ILNT-----GQGEPKEST-----NTW 276
QY 337 MLAGNVNQPYNYDGFNVVYPSSTATVLAASEALA 371
Db 277 MAFLKEKGISHANWSLSDKAPPETGVSVOAGQGV 311
RESULT 12
US-09-465-519-4
; Sequence 4, Application US/09465519
; Patent No. 6403355
; GENERAL INFORMATION:
; APPLICANT: HAGIHARA, Hiroshi
; APPLICANT: KITAYAMA, Kaori
; APPLICANT: HAYASHI, Yasuhiko
; APPLICANT: IGARASHI, Kazuaki
; APPLICANT: ENDO, Keiji
; APPLICANT: OZAKI, Katsuya
; TITLE OF INVENTION: NOVEL AMYLASES
; CURRENT APPLICATION NUMBER: 2173-0118P
; CURRENT FILING DATE: 1999-12-16
; EARLIER APPLICATION NUMBER: 10-362487 JAPAN
; EARLIER FILING DATE: 1998-12-21
; EARLIER APPLICATION NUMBER: 10-362488 JAPAN
; EARLIER FILING DATE: 1998-12-21
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 4
; LENGTH: 501
; TYPE: PRT
; ORGANISM: Bacillus sp.
US-09-465-519-4
Query Match 5.3%; Score 110.5; DB 4; Length 501;
Best Local Similarity 19.6%; Pred. No. 0.049;
Matches 74; Conservative 50; Mismatches 124; Indels 129; Gaps 21;
QY 20 YRYGCTN-----NYLSYQ---SHADVDDVLAQAQAMNLSVIRTWG--FIDIG 62
Db 196 FRFANTNMNRVDEENGNYDYLLGSNTDFSHPEVQDEL-----KDWGSMFTDEL 244
QY 63 SLDCSVPTIDGNKN--GFYQYWDPTGAPAYNDGPTGLQGLDYAIAASAAHGLRVI 120
Db 245 DLDGY--RLDAIKHIPFWYTSWVRHQRNEADOD-----LFVV 280
QY 121 TNDWK-EFGMDQY---DKW-----YGLPYHDNFYDPTQOAY--KNWNHLLNRVNSIT 170
Db 281 GEYKDDVGALEFYLDENMNMESLFDVPLNYNFRASQSGSYDMRNLIRGSLVEAHPMH 340
QY 171 GVTYKNDPTIFAWELANEPKRCVSGTGLPTSGTCTQATIVNWDQMSAYVKSIDPNHNVSV 230
Db 341 AVTFEON-----HDTQP--GESLESWADMFKPL-AYATIL----- 373
QY 231 GDEGFYIGSTQSGWP-----YNDPSDGV-----NNALLRVKNIDFGTYHLYPNY 277
Db 374 -----TREGYPNVFGYDYGIPNDNISAKKIMIDELDQARQNYATGTOHDYFDHW 424
QY 278 GONADWGTOMIKDHIANAAAIKPTILEE-----FGWOTPDORSVYQVOTWTOTVRTNGEA-- 331
Db 425 DV-----VGMTREGSSSRPNNSGLATIMNSGPGSKMYVGRQNGQGTWTDLTGNGASVT 479
QY 332 -----GW-NFWMLAGNVN 343
Db 480 INGDNWGEFFTNGGSVS 496

```

Db 248 FCD-CNEWMR---KLACSYMSRVVVSEFG 272

RESULT 14
US-09-116-622-6
: Sequence 6, Application US/09116522
: Patent No. 6080567
: GENERAL INFORMATION:
: APPLICANT: Kofod, Lene V.
: APPLICANT: Kauppinen, Markus S.
: APPLICANT: Christgau, Stephan
: APPLICANT: Heidt-Hansen, Hans P.
: APPLICANT: Heldt-Hansen, Hans P.
: APPLICANT: Dalboge, Henrik
: APPLICANT: Andersen, Lene N.
: APPLICANT: Si, Joan Q.
: APPLICANT: Jacobson, Tina
: APPLICANT: Munk, Niels
: APPLICANT: Mullertz, Anette
: TITLE OF INVENTION: ENZYMES WITH XYLANASE ACTIVITY FROM
: TITLE OF INVENTION: ASPERGILLUS ACULEATUS
: NUMBER OF SEQUENCES: 42
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: No. 6080567o No. 6080567th America, Inc.
: STREET: 405 Lexington Avenue, 64th Floor
: CITY: New York
: STATE: New York
: COUNTRY: United States of America
: ZIP: 10174-6401
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentin Release #1.0, Version #1.30
: APPLICATION NUMBER: US/09/116.622
: FILING DATE: 16-July-1998
: CLASSIFICATION:
: ATTORNEY/AGENT INFORMATION:
: NAME: Agriis, Cheryl H.
: REGISTRATION NUMBER: 34,086
: REFERENCE/DOCKET NUMBER: 3954.224-US
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 212-867-0123
: TELEFAX: 212-878-9655
: INFORMATION FOR SEQ ID NO: 6:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 296 amino acids
: TYPE: amino acid
: TOPOLOGY: linear
: MOLECULE TYPE: protein
: US-09-116-622-6

Query Match 5.2%; Score 109; DB 3; Length 296;
Best Local Similarity 22.3%; Pred. No. 0.032;
Matches 60; Conservative 37; Mismatches 96; Indels 76; Gaps 16;

Qy 73 GNKNGFYQYW-DPSTGAPAYNDGPTGLOGLDYAIAS-----AAAGHLRVIVVLT 121
Db 47 GYSNGYIYFSTWDGASGDVEYSGAGGSYVTSWSSANFVGKGNPWSAHD-----ITYS 102
Qy 122 NDWKEFGMDYDKWYG-----LPYH---DNFYTPDPTQOAYKKNVHLLNRVNSITGV 172
Db 103 GSWTSTGNSYLSVYGWTTGTPVLEYIILEDYGEYNPWSAGTYKGSVYS-----DGS 154
Qy 173 TYKNDPTIFANELANEPKRCVSGTLP-----TSCTCTQATIVNVWDQMSAYVK--- 220
Db 155 TY----NIVTATRNAPSIOGTATFTQYWSIROTKRVGGTVTTA-----NHFNAWAKLG 204
Qy 221 -SIDPNHMYSVCDGEFYIGSTGSGHPYNDPSGDVNNALLRVKNIDFGTYHLVYPNWGO 279
Db 205 MNLGTHNYQIVATEGY-----SSGSASITVAERAD--ILLR-----YMLY--LWHR 247

US-08-507-431-6
: Sequence 6, Application US/08507431
: Patent No. 5693518
: GENERAL INFORMATION:
: APPLICANT: Kofod, Lene V.
: APPLICANT: Kauppinen, Markus S.
: APPLICANT: Christgau, Stephan
: APPLICANT: Heidt-Hansen, Hans P.
: APPLICANT: Dalboge, Henrik
: APPLICANT: Andersen, Lene N.
: APPLICANT: Si, Joan Q.
: APPLICANT: Jacobson, Tina
: APPLICANT: Munk, Niels
: APPLICANT: Mullertz, Anette
: TITLE OF INVENTION: ENZYMES WITH XYLANASE ACTIVITY FROM
: TITLE OF INVENTION: ASPERGILLUS ACULEATUS
: NUMBER OF SEQUENCES: 42
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: No. 5693518o No. 5693518th America, Inc.
: STREET: 405 Lexington Avenue, 64th Floor
: CITY: New York
: STATE: New York
: COUNTRY: United States of America
: ZIP: 10174-6401
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentin Release #1.0, Version #1.30
: APPLICATION NUMBER: US/08/507.431
: FILING DATE: 15-FEB-1996
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 60/002,800
: FILING DATE: 25-AUG-1996
: ATTORNEY/AGENT INFORMATION:
: NAME: Harrington, James J.
: REGISTRATION NUMBER: 38,711
: REFERENCE/DOCKET NUMBER: 3954.204-US
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 212-867-0123
: TELEFAX: 212-878-9655
: INFORMATION FOR SEQ ID NO: 6:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 296 amino acids
: TYPE: amino acid
: TOPOLOGY: linear
: MOLECULE TYPE: protein
: US-08-507-431-6

Query Match 5.2%; Score 109; DB 1; Length 296;
Best Local Similarity 22.3%; Pred. No. 0.032;
Matches 60; Conservative 37; Mismatches 96; Indels 76; Gaps 16;

Qy 73 GNKNGFYQYW-DPSTGAPAYNDGPTGLOGLDYAIAS-----AAAGHLRVIVVLT 121
Db 47 GYSNGYIYFSTWDGASGDVEYSGAGGSYVTSWSSANFVGKGNPWSAHD-----ITYS 102
Qy 122 NDWKEFGMDYDKWYG-----LPYH---DNFYTPDPTQOAYKKNVHLLNRVNSITGV 172
Db 103 GSWTSTGNSYLSVYGWTTGTPVLEYIILEDYGEYNPWSAGTYKGSVYS-----DGS 154
Qy 173 TYKNDPTIFANELANEPKRCVSGTLP-----TSCTCTQATIVNVWDQMSAYVK--- 220
Db 155 TY----NIVTATRNAPSIOGTATFTQYWSIROTKRVGGTVTTA-----NHFNAWAKLG 204
Qy 221 -SIDPNHMYSVCDGEFYIGSTGSGHPYNDPSGDVNNALLRVKNIDFGTYHLVYPNWGO 279
Db 205 MNLGTHNYQIVATEGY-----SSGSASITVAERAD--ILLR-----YMLY--LWHR 247
Qy 280 NADWGQTQIKDHANAAAIKGPITILEEFG 308

```


GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 13, 2002, 11:55:11 : Search time 7.27371 Seconds
(without alignments)
776.473 Million cell updates/sec

Title: US-09-917-378-3

Perfect score: 2079

Sequence: 1 APAGFTVASCQGVNLGLPY.....YPSSTATVLALEALAISTG 375

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Scarched: 97044 seqs, 15060890 residues

Total number of hits satisfying chosen parameters: 97044

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Published Applications:AA:*
1: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep:*
2: /cgn2_6/ptodata/2/pubpaa/PTCT_NEW_PUB.pep:*
3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep:*
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep:*
5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep:*
6: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep:*
7: /cgn2_6/ptodata/2/pubpaa/PTCTUS_PUBCOMB.pep:*
8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep:*
9: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep:*
10: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep:*
11: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep:*
12: /cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB.pep:*
13: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep:*
14: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep:*

Pred. NO. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	812	39.1	663	9	US-10-121-032-61
2	812	39.1	680	9	US-10-121-032-25
3	493	23.7	427	9	US-09-850-9828-2
4	179.5	8.6	562	10	US-09-981-9008-5
5	124.5	6.0	711	10	US-09-765-272-82
6	117.5	5.7	956	9	US-10-121-032-63
7	117	5.6	951	9	US-09-924-097-15
8	110.5	5.3	501	9	US-09-918-543-27
9	110.5	5.3	501	9	US-09-918-543-28
10	110.5	5.3	501	10	US-09-971-611-4
11	109	5.2	467	10	US-09-863-5478-1
12	104.5	5.0	501	9	US-09-918-543-25
13	104	5.0	1010	10	US-09-118-276-12
14	97	4.7	553	10	US-09-888-224-2
15	95	4.6	877	10	US-09-881-752A-28
16	93.5	4.5	485	9	US-09-918-543-2
17	93.5	4.5	485	10	US-09-769-864-1
18	93.5	4.5	485	10	US-09-769-864-7
19	93.5	4.5	485	10	US-09-854-346-2

20	93.5	4.5	485	10	US-09-902-188A-1	Sequence 1, Appli
21	93	4.5	1430	10	US-09-740-274-6	Sequence 6, Appli
22	92	4.4	514	10	US-09-769-864-3	Sequence 3, Appli
23	92	4.4	514	10	US-09-902-188A-3	Sequence 3, Appli
24	91.5	4.4	472	10	US-09-745-763-36	Sequence 36, Appli
25	91.5	4.4	472	10	US-09-995-587A-11	Sequence 11, Appli
26	91	4.4	515	9	US-09-918-543-6	Sequence 6, Appli
27	91	4.4	515	10	US-09-854-346-6	Sequence 14, Appli
28	89	4.3	925	9	US-09-924-097-14	Sequence 5, Appli
29	88	4.2	386	10	US-09-739-861A-5	Sequence 5, Appli
30	88	4.2	386	10	US-09-795-583-5	Sequence 5, Appli
31	87	4.2	486	10	US-09-815-242-13455	Sequence 13455, A
32	86.5	4.2	672	10	US-09-815-242-10306	Sequence 10306, A
33	86.5	4.2	1375	10	US-09-740-274-4	Sequence 4, Appli
34	85.5	4.1	654	9	US-09-905-291A-177	Sequence 177, App
35	85.5	4.1	654	10	US-09-909-320-177	Sequence 177, App
36	85.5	4.1	654	10	US-09-909-088B-177	Sequence 177, App
37	85.5	4.1	654	12	US-10-052-586-26	Sequence 26, Appli
38	85.5	4.1	1475	10	US-09-740-274-2	Sequence 2, Appli
39	84.5	4.1	156	10	US-09-738-973-210	Sequence 210, App
40	84.5	4.1	1073	10	US-09-815-242-12361	Sequence 12361, A
41	84.5	4.1	1147	10	US-09-815-242-5468	Sequence 5468, Ap
42	84.5	4.1	5795	10	US-09-815-242-12610	Sequence 12610, A
43	84	4.0	275	10	US-09-060-854B-4	Sequence 4, Appli
44	84	4.0	357	10	US-09-837-235-15	Sequence 15, Appli
45	84	4.0	516	10	US-09-986-676A-2	Sequence 2, Appli

ALIGNMENTS

RESULT 1

US-10-121-032-61

: Sequence 61, Application US/10121032

: Patent No. US20020155550A1

: GENERAL INFORMATION:

: APPLICANT: Bylina, Edward J.

: TITLE OF INVENTION: GLYCOSIDASE ENZYMES

: NUMBER OF SEQUENCES: 72

: CORRESPONDENCE ADDRES:

: ADDRESSEE: Gray Cary Ware & Freidenrich LLP

: STREET: 4365 Executive Drive, Suite 1600

: CITY: San Diego

: STATE: CA

: COUNTRY: USA

: ZIP: 92121

: COMPUTER READABLE FORM:

: MEDIUM TYPE: Diskette

: COMPUTER: IBM Compatible

: OPERATING SYSTEM: Windows95

: SOFTWARE: FastSeq for Windows Version 2.0

: CURRENT APPLICATION DATA:

: APPLICATION NUMBER: US/10/121,032

: FILING DATE: 09-Apr-2002

: CLASSIFICATION: <Unknown>

: PRIOR APPLICATION DATA:

: APPLICATION NUMBER: US/09/134,078

: FILING DATE: 13-AUG-1998

: APPLICATION NUMBER: 08/949,026

: FILING DATE: 10-OCT-1997

: APPLICATION NUMBER: 60/056,916

: FILING DATE: 06-DEC-1996

: ATTORNEY/AGENT INFORMATION:

: NAME: Halie, Lisa A.

: REGISTRATION NUMBER: 38,347

: REFERENCE/DOCKET NUMBER: 09010/024002

: TELECOMMUNICATION INFORMATION:

: TELEPHONE: 858/677-1456

: TELEFAX: 858/677-1465

: INFORMATION FOR SEQ ID NO: 61:

: SEQUENCE CHARACTERISTICS:

: LENGTH: 663 amino acids

: TYPE: amino acid

```

:
: TOPOLOGY: linear
: MOLECULE TYPE: protein
: FRAGMENT TYPE: internal
: SEQUENCE DESCRIPTION: SEQ ID NO: 61:
US-10-121-032-61

Query Match 39.18; Score 812; DB 9; Length 663;
Best Local Similarity 42.88; Pred. No. 7.2e-62;
Matches 169; Conservative 61; Mismatches 121; Indels 44; Gaps 13:

QY 5 FVTASGGQFVLUNGLPYRYGCTNNTYLSYQSHADVDVLLAKAQAAMNLSVIRTWGFTIDIGSL 64
  || || || || || || || || || || || || || || || || || || || || || || ||
DB 17 FVKVENGKFAALNGKREFRIGSNVNYMHYKSGMIDSVLESARDMCIKVLRIWGF----L 71

QY 65 DGSVPTIDGNKNGFYFQWDRSTCAPYNDG----PTCQLQGLDYAIAASAAHGLRIVVL 120
  || || || || || || || || || || || || || || || || || || || || || || ||
DB 72 DGSYCRDKN-----TYMHPEPGVFCVPEGISNAQSGFERLDTYVAKAKELGKLIVL 125

QY 121 TNDWKEFGMDQYDKWYGLPLYHNFYDTPRTOQAQYKNNVNILLNRVNSITGVYKNDPTI 180
  || || || || || || || || || || || || || || || || || || || || || || ||
DB 126 VNNWDDFCGMNQYVRVFGCTHDDFYRDEKIKEEYKVVSVFLNVHVNVTYTCVPRREPTI 185

QY 181 FAWELANEPKRCVSGTLPSTCTCTQATIVNVWDOMSAYKSIDPNHMYSVGDEGF---YI 237
  || || || || || || || || || || || || || || || || || || || || || || ||
DB 186 MAWELANEPKRC-----ETDKSGN---TLVENWVKEMSSYIKSLDPNHLVAVGDEGFFSNYE 237

QY 238 GSTGSG---WPYNDPSGDVNNALLRVKNIDFGTYHLYPNWYGN-----ADWCTOWIKD 290
  || || || || || || || || || || || || || || || || || || || || || || ||
DB 238 GFKPYGGEAEWAYNGWS-GVDWKKLLSIETVDFGTFLYPSHWGVSPENYQWGAKWIED 296

QY 291 HIAANAAIGKPTILEEFG--WOTP-DRDSVYQTWTQTVTRTNGEAGWNFMWLAGNVNGOP- 346
  || || || || || || || || || || || || || || || || || || || || || || ||
DB 297 HIKIAKEIGKPVVLEEYCIKPSAPNRTAIRLVNLDLYDLGGDAMPWMLAGICEGSDR 356

QY 347 -----YPNYDCGFNYVPSSTATVLASE--ALAIATG 375
  || || || || || || || || || || || || || || || || || || || || || || ||
DB 357 DERGYPDYDGFVRVNDSDPEALIREYAKLNFNG 391

RESULT 2
US-10-121-032-25
: Sequence 25, Application US/10121032
: Patent No. US20020155550A1
: GENERAL INFORMATION:
: APPLICANT: Bylina, Edward J.
: TITLE OF INVENTION: GLYCOSIDASE ENZYMES
: NUMBER OF SEQUENCES: 72
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Gray Cary Ware & Freidenrich LLP
: STREET: 4365 Executive Drive, Suite 1600
: CITY: San Diego
: STATE: CA
: COUNTRY: USA
: ZIP: 92121
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Diskette
: COMPUTER: IBM Compatible
: OPERATING SYSTEM: Windows95
: SOFTWARE: FASTSEQ for Windows Version 2.0
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/10121.032
: FILING DATE: 09-Apr-2002
: CLASSIFICATION: <unknown>
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US/09/134.078
: FILING DATE: 13-AUG-1998
: APPLICATION NUMBER: 08/949,026
: FILING DATE: 10-OCT-1997
: APPLICATION NUMBER: 60/056,916
: FILING DATE: 06-DEC-1996
: ATTORNEY/AGENT INFORMATION:
: NAME: Haile, Lisa A.
: REGISTRATION NUMBER: 38,347

```

```

: REFERENCE/DOCKET NUMBER: 09010/024002
: TELECOMMUNICATION INFORMATION:
:   TELEPHONE: 858/677-1456
:   TELEFAX: 858/677-1465
: INFORMATION FOR SEQ ID NO: 25:
: SEQUENCE CHARACTERISTICS:
:   LENGTH: 680 amino acids
:   TYPE: amino acid
:   TOPOLOGY: linear
: MOLECULE TYPE: protein
: FRAGMENT TYPE: internal
: SEQUENCE DESCRIPTION: SEQ ID NO: 25:
US-10-121-032-25

Query Match          39.1%; Score 812; DB 9; Length 680;
Best Local Similarity 42.8%; Pred. No. 7.5e-62;
Matches 169; Conservative 61; Mismatches 121; Indels 44; Gaps 15;

QY    5 FVTASGCGFVLNGLPYRYGGTNNYYLSQSADVDDVLAKAQAANLNSVIRTWGFIDIGSL 64
      || |::||| :::|::||| :::|::||| :::|::||| :::|::||| :::|::||| :
Db    34 FVKVENGKEALNGEKFRIFGSNYYMHYKSGMIDSVLESARDNGIKVLRTWGP-----L 88

QY    65 DGSVPTIDGNKNRGFYFOVDWPSTCAPANDG----PTGLOGLDYAIASAAGLRVIVLI 120
      || |::||| :::|::||| :::|::||| :::|::||| :::|::||| :::|::|||
Db    89 DGESYCRDKN-----TYMHPEPCVFCGPEGISNAOSGFERLDYTVAKAKELGIKLIVL 142

QY    121 TNDWKKEFGMDQDKWGLPYLPHDNFYDTPRTOQAYKWNHLLNRVNSITGVITYKNDIPTI 180
      || |::||| :::|::||| :::|::||| :::|::||| :::|::||| :::|::|||
Db    143 VNNWDFEGGMNYVRVFEGTHDDFDIKEYEKKYVSFLVNHHVNTYTGVPRKEPTI 202

QY    181 PAWELANEPCRCVSGTLPTSGTGCTCOATIINWVDQMSAVKSIDPNHMVSVSDECF---YI 237
      || |::||| :::|::||| :::|::||| :::|::||| :::|::||| :::|::|||
Db    203 MAWELANEPRC-----ETDKSGN----TLVEWVKEMSSYIKSLDPNHLVAVGDGEFFSNYE 254

QY    238 GSTOGSG---WPYNDRPSGDVDNNALLRVKNIIDFTGYHLYPNYWGON-----ADMGTOWIKD 290
      | |::||| :::|::||| :::|::||| :::|::||| :::|::||| :::|::|||
Db    255 GFKPYGGAEWAYNGWS-GVDMKKLLSIETVDFCTFHPYSHMGVSPENYAQMCAKWIED 313

QY    291 HIANAAIGKPTILEECG-WQTP-DROSVYOTWTQVTRTNGEAGHNFWMLAGNVGNCP- 346
      || |::||| :::|::||| :::|::||| :::|::||| :::|::||| :::|::|||
Db    314 HKIAKEIGKPVLEEYGIKPSAPVNTATIRLMNDILVYDLGGDGAMFWMLAGIGEGR 373

QY    347 -----YPNYDGFNYYYSPSSATVLASE--ALAISTG 375
      || |::||| :::|::||| :::|::||| :::|::||| :::|::||| :::|::|||
Db    374 DERCYYPDYDCFRIVNDDSPAEELIREYAKLFNTG 408

RESULT 3
us-09-850-982B-2
: Sequence 2, Application US/09850982B
: Patent No. US20020166145A1
: GENERAL INFORMATION:
: APPLICANT: Nestec S.A.
: TITLE OF INVENTION: COFFEE MANNAHAVE
: FILE REFERENCE: 88265-4025
: CURRENT APPLICATION NUMBER: US/09/850,982B
: CURRENT FILING DATE: 2001-05-08
: NUMBER OF SEQ ID NOS: 12
: SOFTWARE: Patentin version 3.1
: SEQ ID NO 2
: LENGTH: 427
: TYPE: PRT
: ORGANISM: Coffea arabica
US-09-850-982B-2

Query Match          23.7%; Score 493; DB 9; Length 427;
Best Local Similarity 32.1%; Pred. No. 8.2e-35;
Matches 125; Conservative 62; Mismatches 137; Indels 66; Gaps 16;

QY    5 FVTASGCGFVLNGLPYRYGGTNNYYLSYS-----QSHADVDDVLAKAQAMNLNSVIRTWGFI 59
      || |::||| :::|::||| :::|::||| :::|::||| :::|::||| :::|::|||
Db    38 FIOTRGTRFVLGGYPFFNFNGFNFSNMWHVAAPERHKISNWFEAAATGLTCVRTWAFS 97

```

```

: REFERENCE/DOCKET NUMBER: 09010/024002
: TELECOMMUNICATION INFORMATION:
:   TELEPHONE: 858/677-1456
:   TELEFAX: 858/677-1465
: INFORMATION FOR SEQ ID NO: 25:
: SEQUENCE CHARACTERISTICS:
:   LENGTH: 680 amino acids
:   TYPE: amino acid
:   TOPOLOGY: linear
: MOLECULE TYPE: protein
: FRAGMENT TYPE: internal
: SEQUENCE DESCRIPTION: SEQ ID NO: 25:
US-10-121-032-25

Query Match          39.1%; Score 812; DB 9; Length 680;
Best Local Similarity 42.8%; Pred. No. 7.5e-62;
Matches 169; Conservative 61; Mismatches 121; Indels 44; Gaps 15;

QY      5 FVTASGCGFVLNGLPYRYGGTNNYYLSQSADVDDVLAKAQAANLNSVIRTWGFIDIGSL 64
       || |::||| ::::|::||| ::::|::||| ::::|::||| ::::|::||| ::::|::|||
DB     34 FVKVENGKEALNGEKFRIFGSNYYMHYKSGMIDSVLESARDNGIKVLRTWGP-----L 88

QY     65 DGSVPTIDGNKNRGFYFOVWPSTCAPAYNDG----PTGLOGLDYAIASAAGLRVIVLI 120
       || |::||| ::::|::||| ::::|::||| ::::|::||| ::::|::||| ::::|::|||
DB     89 DGESYCRDN-----TYMHPEPCVFCPEGISNAOSGFERLDYTVAKAKELGIKLIVL 142

QY    121 TNDWKFGGMDQDKWGLPYLPHDFNYDTPRTOQAYKWNHLLNRVNSITGVITYKNDIPTI 180
       || |::||| ::::|::||| ::::|::||| ::::|::||| ::::|::||| ::::|::|||
DB    143 VNNWDFEGGMNQYVRVFEGTHDDFDIKESYKKYSVFLVNHHVNTYTGVPRKEPTI 202

QY    181 PAWELANEPCRCVSGTLPTSGTGCTCAATIVNWVDQMSAVKSIDPNHMVSVCDEGF---YI 237
       || |::||| ::::|::||| ::::|::||| ::::|::||| ::::|::||| ::::|::|||
DB    203 MAWELANEPRC-----ETDKSGN----TLVEWVKEMSSYIKSLDPNHLVAVGDGEFFSNYE 254

QY    238 GSTOGSG---WPYNDRPSGDVDNNALLRVKNIIDFTGYHYLPNYMGON----ADMGTOWIKD 290
       || |::||| ::::|::||| ::::|::||| ::::|::||| ::::|::||| ::::|::|||
DB    255 GFKPYGGAEWAYNGWS-GVDWKLLSLETVDFTGTFHPYSHMGVSPENYAQMCAKWIED 313

QY    291 HIANAAIGKPTILEECG-WQTP-DROSVYOTWTQVTRTNGEAGHWFMWLAGNVGNCP-- 346
       || |::||| ::::|::||| ::::|::||| ::::|::||| ::::|::||| ::::|::|||
DB    314 HKIAKEIGKPVLEEYGIKPSAPVNTAIRUMNDILVYDLGGDGAMFWMLAGIGEGR 373

QY    347 -----YPNYDGFNYVPSSSTATVLASE-ALAISTG 375
       || |::||| ::::|::||| ::::|::||| ::::|::||| ::::|::||| ::::|::|||
DB    374 DERCYYPDYDCFRIVNDDSPAEELIREYAKLFNTG 408


RESULT 3
us-09-850-982B-2
: Sequence 2, Application US/09850982B
: Patent No. US20020166145A1
: GENERAL INFORMATION:
: APPLICANT: Nestec S.A.
: TITLE OF INVENTION: COFFEE MANNAHAVE
: FILE REFERENCE: 88265-4025
: CURRENT APPLICATION NUMBER: US/09/850,982B
: CURRENT FILING DATE: 2001-05-08
: NUMBER OF SEQ ID NOS: 12
: SOFTWARE: Patentin version 3.1
: SEQ ID NO 2
: LENGTH: 427
: TYPE: PRT
: ORGANISM: Coffea arabica
US-09-850-982B-2

Query Match          23.7%; Score 493; DB 9; Length 427;
Best Local Similarity 32.1%; Pred. No. 8.2e-35;
Matches 125; Conservative 62; Mismatches 137; Indels 66; Gaps 16;

QY      5 FVTASGCGFVLNGLPYRYGGTNNYYLSY-----QSHADVDDVLAKAQAMNLNSVIRTWGFI 59
       || |::||| ::::|::||| ::::|::||| ::::|::||| ::::|::||| ::::|::|||
DB     38 FIOTRGTRFVLGGYPFFNFGENFSNMWHVAAPERHKISNWFREAAATGLTCVRTWAFS 97

```


Db 589 YKALRAKHPKWLIIYGSETSSATRTKSYRPERELKHSNCPERNYEQSDYGNDRVCGWGT 648
Qy 281 ADWGTQWIKDHIANAAA-----IGKPTILEEFQWQTPDRDSVYQVWTQVTRTNGEAG 332
Db 649 ATASWTFDRDAGYAGQFIWTCDTDICEPT-----PWHNQNOTPVKSSYFEGIVDTAGIPK 703
Qy 333 WNFWM 337
Db 704 HIFYL 708

RESULT 6
US-10-121-032-63
; Sequence 63, Application US/10121032
; Patent No. US20020155550A1
; GENERAL INFORMATION:
; APPLICANT: Bylina, Edward J.
; TITLE OF INVENTION: GLYCOSIDASE ENZYMES
; NUMBER OF SEQUENCES: 72
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Gray Cary Ware & Freidenrich LLP
; STREET: 4365 Executive Drive, Suite 1600
; CITY: San Diego
; STATE: CA
; COUNTRY: USA
; ZIP: 92121
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: Windows95
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/121,032
; FILING DATE: 09-AUG-2002
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/09/134,078
; FILING DATE: 13-AUG-1998
; APPLICATION NUMBER: 08/949,026
; FILING DATE: 10-OCT-1997
; APPLICATION NUMBER: 60/056,916
; FILING DATE: 06-DEC-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Halle, Lisa A.
; REGISTRATION NUMBER: 38,347
; REFERENCE/DOCKET NUMBER: 09010/024002
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 858/677-1456
; TELEFAX: 858/677-1465
; INFORMATION FOR SEQ ID NO: 63:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 956 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; FRAGMENT TYPE: Internal
; SEQUENCE DESCRIPTION: SEQ ID NO: 63:
US-10-121-032-63

Query Match 5.7%; Score 117.5; DB 9; Length 956;
Best Local Similarity 19.2%; Pred. No. 0.031;
Matches 76; Conservative 52; Mismatches 110; Indels 157; Gaps 19;
Qy 3 AGFVTASGGQFVLNGLPY-RYGGTNNYLSYQSHADVDVLAQAAAMNLSVIRTWGFI 61
Db 48 SGEITSLAG-----NSLFWNAGTDSIFY-----NATVDFL--AENWNSLIR-----IAM 92
Qy 62 GSLDGSVPTIDGNKNGFYQYWDPSGTGAPAYNDGPTGLQGLD-LDYAIAASAAHGLRIVVYL 120
Db 93 GYKEN-----WD-----GGNGYIDSPQEQEAKIRKVIDAALANGIYVII-- 131
Qy 121 TNDWKEFGMDQYDKWGLPYHDNFYDTPRTOQAYKNWNNHLLNRVNSITGVYTK--NDP 178

Db 132 --DMH-----THEAEVLTDEAV-----DFFTRMADLYCDTNPVNYEIVNEP 170
Qy 179 TIFAWELANEPKRCVSGTLPSTCTCTQATIVNVDMSAYVKSIDPNIMVSVGDEGFIYIC 238
Db 171 IYQSWPV-----IKNYAEQVIAGIRSKDPDNLIIVGTSNYSQO 208
Qy 239 STOGSGMPYND-----PSD-----CVDNNALLRVKNIDFGTYHLYPNY 276
Db 209 VDVASADPISDTNWAVYTLHFYAAFNPNDNLNRNVAQTALDNNVALFV----- 254
Qy 277 WGNADWGTQWIKDHIANAAAIGKPTILEEFQWQTPDRDSVYQVWTQVTRTNGEAGNFW 336
Db 255 ----TEWGT-----LNT-----GQEPDKEST-----NTW 276
Qy 337 MLAGNVNGQYPNVDGFNVYSPSTATVLAESALA 371
Db 277 MAFLKEKGISHANWSLSKAPPETGSVVOAGOGVS 311

RESULT 7
US-09-924-097-15
; Sequence 15, Application US/09924097
; Patent No. US20020156240A1
; GENERAL INFORMATION:
; APPLICANT: TOMONO, Jun
; APPLICANT: NOMURA, Yoshiko
; APPLICANT: SAGAWA, Hiroaki
; APPLICANT: SAKAI, Takeshi
; APPLICANT: KATO, Ikunoshin
; TITLE OF INVENTION: ALPHA-AGARASE AND PROCESS FOR PRODUCING THE SAME
; FILE REFERENCE: TOMONO-1
; CURRENT APPLICATION NUMBER: US/09/924,097
; CURRENT FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: JP00/00966
; PRIOR FILING DATE: 2000-02-21
; PRIOR APPLICATION NUMBER: 11-44890
; PRIOR FILING DATE: 1999-02-23
; PRIOR APPLICATION NUMBER: 11-198852
; PRIOR FILING DATE: 1999-07-13
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 15
; LENGTH: 951
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Amino acid sequence of agarase 4-3
US-09-924-097-15
Query Match 5.6%; Score 117; DB 9; Length 951;
Best Local Similarity 20.1%; Pred. No. 0.033;
Matches 74; Conservative 45; Mismatches 113; Indels 136; Gaps 21;
Qy 4 GFVTASGG---QFVLNGLPYRGCTNNYLL-----SYQSHADVDVLA 43
Db 209 GFVAAAGATNVSWVTNG---DFGDYNNIYLEAGTYRSFITVATGSDGSGYCARIDLDGSPA 265
Qy 44 KAAMNLSVIRTWGFTIDGSLDGSVPTIDGNKNGFYQYWDPSGTGAPAYNDGPTGLQGLD 103
Db 266 -----SMGYFD--STGG-----WETPEEVELY-----GGD 288
Qy 104 YATASAAHGLRIVVLTNDWKEFGMDQYDKWGLPYHDNFYDTPRTOQAYK-----NW 158
Db 289 FVYETSGHTFLRVEAIGGSDW-----QWSG-----DNVRLTKVGDASVKPSPLYNP 334
Qy 159 VNHLLNRVNS-ITGVTYKNDPTIFANELANEPKRCVSGTLPSTCTCTQATIVNVDMSA 217
Db 335 KDHTVTEIEGPVVGLEFLKKPV-----QVPTANRLKSDV--WYTY--- 373
Qy 218 YVKSIDPNIMVSVGDEGFIYQSGWPNYNDPSDGVNNALLRVKNIDFGTYHLYPNY 277
Db 374 -----PONNELQGFQNF--GAT-GSFWGHPPPEENFYDOTTI-----IDW--TOLVQNYQ 417


```
Best Local Similarity 19.6%; Pred. No. 0.054;
Matches 74; Conservative 50; Mismatches 124; Indels 129; Gaps 21;

QY 20 YRYGGTN-----NYILSYQ---SHADVDDVLAKAQAAMNLSVIRTWG--FIDIG 62
DB 196 FRFANTNNWRYDEENGNYDYLLGSNIDFSHPEVODEL-----KDWGSWFTDEL 244
QY 63 SLDGSVPTIDGKN--GFYFYQWDPSTGAPAYNDGPTGLOGLDYAIASAAAHCLRVIVVL 120
DB 245 DLDGY--RLDAIKH1PFWYTSDWVRHORNEADQD-----LFFV 280
QY 121 TNDWK-EFGGMDQY---DKW-----YGLPYHDNFEYTDPTQOAY--KNWVHLLNRVNSIT 170
DB 281 GEYWKDDVGALFYLDEMNMWMSLFDVPLNYNFRASQCGSYDMRNILRGSLVEAHPH 340
QY 171 GVTYKNDPTIFAWELANPRCVSGTLPSTSGTCTQATIVNWVDQMSAYVKSIDPNHMSV 230
DB 341 AVTFVDN-----HDTQP--GESLESWADWFKPL-AYATIL----- 373
QY 231 GDEGFYIGTQSGWP-----YNDPSCVD-----NNALLRVKNIDFGTYHYLPNYW 277
DB 374 -----TREGGYPNVFYGDYIGIPNDNISAKKMDIDELLDARQNYAYGTOHDFDHW 424
QY 278 GQADWGTOWIKDHIANAALCKPTILEE-----FGWQTPDRDSVYQVOTWQ-----TVR 326
DB 425 DV-----VGTWREGSSSRPNSGLATIMSNPGCGSKMYYVGRONAGOTWTDLTGNCASVT 479
QY 332 ----GW-NFWMLAGNVN 343
DB 480 INGDGNGEFTNGGSVS 496

RESULT 11
US-09-863-547B-1
; Sequence 1, Application US/09863547B
; Patent No. US20020128166A1
; GENERAL INFORMATION:
; APPLICANT: Henkel KGaA
; APPLICANT: Hermanus, Lentus B. M.
; APPLICANT: Van Beckhoven, Rudolf F. W. C.
; APPLICANT: Maurer, Karl-Helmut
; APPLICANT: Kottwitz, Beatrix
; APPLICANT: Weiss, Albrecht
; APPLICANT: Van Solingen, Pieter
; TITLE OF INVENTION: Detergents Comprising Cellulases
; FILE REFERENCE: H 1920 A
; CURRENT APPLICATION NUMBER: US/09/863,547B
; CURRENT FILING DATE: 2002-02-13
; PRIOR APPLICATION NUMBER: US 08/945,574
; PRIOR FILING DATE: 1998-02-27
; PRIOR APPLICATION NUMBER: PCT/EP96/01755
; PRIOR FILING DATE: 1996-04-26
; PRIOR APPLICATION NUMBER: US 08/614,115
; PRIOR FILING DATE: 1996-03-12
; PRIOR APPLICATION NUMBER: EP 95201115.3
; PRIOR FILING DATE: 1995-04-28
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
; LENGTH: 467
; TYPE: PRT
; ORGANISM: Bacillus sp. CBS 670.93
US-09-863-547B-1

Query Match 5.2%; Score 109; DB 10; Length 467;
Best Local Similarity 22.3%; Pred. No. 0.066;
Matches 59; Conservative 31; Mismatches 96; Indels 78; Gaps 16;

QY 87 TGAPAYNDGPTGLOGLDYAIAASAAHCLRVIVVLNDEMKEFGMDQYDKWYGLPYHDNPFY 146
DB 93 TSSGGYIDDPVSKVKETVEAIDLGIYVII-----DWHILS-----DN-- 132
QY 147 TDPRTQOAYKNWVHLLNRVNSITGVTYKNDPTIFAWELANPRCVSGTLPSTSGTCTQOA 206
```

```
DB 133 -DP---NIYKEAKDFDEMSELYG-DYPN-----VIVEIANEPN--GSD----- 170
QY 207 TIVNWVDQMSAYVKS-----DPNIMVSVGDEGFYIGTQSGWGPYPNDPSGCVDDNNAL 259
DB 171 --VTWQNIKPAAEEVIPVIRONDNNIVIV-----CTGTWSDOVHHAADNQ-- 215
QY 260 LRVKNIDFGTYHYLPNYWGNADWGTOWIKDHIANAALCKPTILEEFGWQTPDRD-SVY 318
DB 216 LADPNVY-AFHFYAGTHGON-----LRDQVDYALDQGAALFVSEWGTSAATGDCGVF 267
QY 319 ----QWTQTQVTRTNGEAGWNFWML 338
DB 268 LDEAQVWIDFM-DERNLSWANWSL 290

RESULT 12
US-09-918-543-25
; Sequence 25, Application US/09918543
; Patent No. US20020155574A1
; GENERAL INFORMATION:
; APPLICANT: No. US20020155574A1ozymes A/S
; APPLICANT: Thisted, Thomas
; APPLICANT: Kjaerulff, Soren
; APPLICANT: Andersen, Carsten
; APPLICANT: Fuglsang, Claus Crone
; TITLE OF INVENTION: Alpha-amylase mutants with altered properties
; FILE REFERENCE: 10062.200-US
; CURRENT APPLICATION NUMBER: US/09/918,543
; CURRENT FILING DATE: 2001-07-31
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 25
; LENGTH: 501
; TYPE: PRT
; ORGANISM: Bacillus sp.
US-09-918-543-25
```

```
Query Match 5.0%; Score 104.5; DB 9; Length 501;
Best Local Similarity 19.1%; Pred. No. 0.17;
Matches 72; Conservative 52; Mismatches 124; Indels 129; Gaps 20;

QY 20 YRYGGTN-----NYILSYQ---SHADVDDVLAKAQAAMNLSVIRTWG--FIDIG 62
DB 196 FRFANTNNWRYDEENGNYDYLLGSNIDFSHPEVOEL-----KDWGSWFTDEL 244
QY 63 SLDGSVPTIDGKN--GFYFYQWDPSTGAPAYNDGPTGLOGLDYAIAASAAHCLRVIVVL 120
DB 245 DLDGY--RLDAIKH1PFWYTSDWVRHORSEADQD-----LFFV 280
QY 121 TNDWK-EFGGMDQY---DKW-----YGLPYHDNFEYTDPTQOAY--KNWVHLLNRVNSIT 170
DB 281 GEYWKDDVGALFYLDEMNMWMSLFDVPLNYNFRASQCGSYDMRNILRGSLVEAHPH 340
QY 171 GVTYKNDPTIFAWELANPRCVSGTLPSTSGTCTQATIVNWVDQMSAYVKSIDPNHMSV 230
DB 341 AVTFVDN-----HDTQP--GESLESWADWFKPL-AYATIL----- 373
QY 231 GDEGFYIGTQSGWP-----YNDPSCVD-----NNALLRVKNIDFGTYHYLPNYW 277
DB 374 -----TREGGYPNVFYGDYIGIPNDNISAKKMDIDELLDARQNYAYGTOHDFDHW 424
QY 278 GQADWGTOWIKDHIANAALCKPTILEE-----FGWQTPDRDSVYQVOTWQ-----TVR 326
DB 425 DI-----VGTWREGSSSRPNSGLATIMSNPGCGSKMYYVGRONAGOTWTDLTGNCIAASVT 479
QY 327 TNGEAGWNFWMLAGNVN 343
DB 480 INGDGNGEFTNGGSVS 496

RESULT 13
US-09-118-276-12
```

Sequence 12, Application US/09118276
Patent No. US20010011381A1
GENERAL INFORMATION:
APPLICANT: BABYCHUK, ELENA;
APPLICANT: KUSHNIR, SERGEI;
APPLICANT: DE BLOCK, MARC;
APPLICANT: INZE, DIRK
TITLE OF INVENTION: METHODS AND MEANS TO MODULATE PROGRAMMED
CELL DEATH IN EUKARYOTIC CELLS
NUMBER OF SEQUENCES: 21
CORRESPONDENCE ADDRESS:
ADDRESSEE: SIXBEY, FRIEDMAN, LEEDOM, & FERGUSON
STREET: 8180 GREENSBORO DRIVE, SUITE 800
CITY: MCLEAN,
STATE: VIRGINIA
COUNTRY: USA
ZIP: 22102
COMPUTER READABLE FORM:
MEDIUM TYPE: 3-1/2" DISKETTE
COMPUTER: IBM-COMPATIBLE
OPERATING SYSTEM: MS-DOS
SOFTWARE: ASCII
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/118,276
FILING DATE: 17-JUL-1998
PRIOR APPLICATION DATA: NONE
ATTORNEY/AGENT INFORMATION:
NAME: SCHULMAN, ROBERT M.; SALEN, KENNETH H.
REGISTRATION NUMBER: 31,196; 43,077
REFERENCE/DOCKET NUMBER: 6201-0003
TELEPHONE: (703) 790-9110
TELEFAX: (703) 883-0370
INFORMATION FOR SEQ ID NO: 12:
SEQUENCE CHARACTERISTICS:
LENGTH: 1010 RESIDUES
TYPE: AMINO ACID
STRANDEDNESS: SINGLE
TOPOLOGY: LINEAR
US-09-118-276-12

Query Match 5.0%; Score 104; DB 10; Length 1010;
Best Local Similarity 19.3%; Pred. No. 0.47;
Matches 70; Conservative 49; Mismatches 127; Indels 116; Caps 16;
QY 6 VTASGGQFVLNGLPYRYGCTNNYLSYQSHADVDVLAQAQANLNVIRTWGFDIGSLD 65
DB 681 VAVKGEQFLINHKPP-----YFTGGRHEDAD--LRGKGFNVLVHDHALMD----- 726
QY 66 GSVPTIDGNKNGFYQYWDPTGAPAYNDG--PTGLQGLDYAIASAAAGLRLVIVLTND 123
DB 727 -----W-----ICANSYRTSHYPYAEMLDW---ADEHG---IVVIDE 758
QY 124 WKEFGGMDQYKYGGLPYH-----DNFYTD-----RTQQAYKNVHLLNRVNSITGVTY 174
DB 759 TAAVG-----FNLSLGIGFEAGNPKELYSEAVNGETQQAHLQAIKELIAR-----D 806
QY 175 KNDPTIFAMELANEPRCVSGTLPTSCCTQATIVNVDMSAYVKSIDPNHNVSVGDEG 234
DB 807 KNIHPSVVMMSIANEPDTRQAGREYFAPLAET-----RKLDPTRTITCVNVM 854
QY 235 FYIGSTQSGWPNDSQGVNDVNALLRVKNIDFGTYHLYPNYWGQNDWGTQ---WIKDH 291
DB 855 FCAHPT-----DTISDLFDVLCILNR-----YGVVYVQSCDLETAEKVLEKEL 896
QY 292 IANAAAIKPTILEEFG-----NQTPDRDSVYQTWGTQVTRTNGEAGH 333
DB 897 LAWQEKLHPITITGYVDTLAGLHSMYTDMMSEYOCALMDMYHRVDFRVSVAVGEQVW 956
QY 334 NF 335
DB 957 NF 958

RESULT 14
US-09-888-224-2
Sequence 2, Application US/09888224
Patent No. US20020120118A1
GENERAL INFORMATION:
APPLICANT: Lam, D. et al.
TITLE OF INVENTION: Endoglucanases
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESSEE: CARELLA, BYRNE, BAIN, GILFILLAN,
STREET: 6 BECKER FARM ROAD
CITY: ROSELAND
STATE: NEW JERSEY
COUNTRY: USA
ZIP: 07068
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5 INCH DISKETTE
COMPUTER: IBM PS/2
OPERATING SYSTEM: MS-DOS
SOFTWARE: WORD PERFECT 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/888,224
FILING DATE:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/09/430,669
FILING DATE: 28-OCT-1999
APPLICATION NUMBER: US/08/651,572
FILING DATE: 22-MAY-1996
ATTORNEY/AGENT INFORMATION:
NAME: Charles J. Herron
REGISTRATION NUMBER: 28,019
REFERENCE/DOCKET NUMBER: 331400-48
TELEPHONE: 201-994-1700
TELEFAX: 201-994-1744
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 553 AMINO ACIDS
TYPE: AMINO ACID
STRANDEDNESS: Unknown
TOPOLOGY: LINEAR
MOLECULE TYPE: PROTEIN
US-09-888-224-2

Query Match 4.7%; Score 97; DB 10; Length 553;
Best Local Similarity 18.4%; Pred. No. 0.87;
Matches 83; Conservative 49; Mismatches 142; Indels 178; Caps 24;
QY 25 TNNYLSYQSHADVDVLAQAQANLNVIRTWGFDIGSLDGSVPT-IDGNKNGFYQYW 83
DB 24 TPNVYVHGLSNRMEDMLQIKSLGFNAIRL-PECTOSVKPGTMTPTAIDYAKN----- 75
QY 84 DPSTGAPAYNDGPTGQLDYA-----IASAAAGLRLVIVLTNDWKEFGMDQYDKWY 137
DB 76 -----PDLOGLDSVOIMEKLIKAGDLGIFVL----- 102
QY 138 GLPYHD---NF-----YTDPTQOAYKN-WYN--HLLNRVNSITGVTKNDPTIFAMELA 186
DB 103 -LDYHRIHCNFIPLWYTFDSFSEQDYINTWVEVAQREKYNVWIGADLKNKP-----HSS 157
QY 187 NEPRCVSGTLPSTGCTQATIVNW---VDMSAYVKSIDPNHNVSVGDSGFYIGSTQG- 242
DB 158 PAPAAYTDGSGATWGMGNAT--DWNLAERICRAILEVAPQWYIFVEGTQFTTPEIDGR 215
QY 243 -----SGW-----PYNDPSDGVNDNALLRVK 263
DB 216 YKWHNNAWCGNLMGVRYKYPVNLPRDKVYVSPQYGVSEYDQYFDPGEGFPDN----- 269
QY 264 NIDFGTYHLYPNYWGQNDWGTQWIKDHIANAAAIKPTILEEFG----- 308
DB 270 -----LPEIWHH--FG--YVK-----LDLGYPVVIGEGGKYGHGGDRDVTWQ 310

Search completed: November 13, 2002, 12:08:33
Job time : 10.2737 secs

QY 309 -----WOTPDR--DSVYOTWTVRTNGEAG-----W-NFW-----MLAGNV 342
Db 311 NKIIDWMIQKCDFFYWSNP---NSGDTGGILKDDWTTIWEDKYNLKLRLMDSCSNA 367
QY 343 NGOPYNYDCFNVPYPSSTATVLALEALIST 374
Db 368 TAPSVPTTTTSTPTPTTTTSTPTTTTQT 399

RESULT 15

US-09-881-752A-28
: Sequence 28, Application US/09881752A
: Patent No. US20020115078A1
: GENERAL INFORMATION:
: APPLICANT: Kleanthous, Harold
: APPLICANT: Al-Garawi, Amal
: APPLICANT: Miller, Charles
: APPLICANT: Tomb, Jean-Francois
: APPLICANT: Oomen, Raymond P.
: TITLE OF INVENTION: Identification of Polynucleotides
: TITLE OF INVENTION: Encoding No. US20020115078A1 Helicobacter Polypeptides in the
: FILE REFERENCE: 06132/041002
: CURRENT APPLICATION NUMBER: US/09/881,752A
: PRIOR FILING DATE: 2001-06-15
: PRIOR APPLICATION NUMBER: US 08/833,457
: PRIOR FILING DATE: 1997-04-01
: NUMBER OF SEQ ID NOS: 370
: SOFTWARE: FastSeq for Windows Version 4.0
: SEQ ID NO 28
: LENGTH: 877
: TYPE: PRT
: ORGANISM: Helicobacter pylori
US-09-881-752A-28

Query Match 4.6%; Score 95; DB 10; Length 877;
Best Local Similarity 20.9%; Pred. No. 2.3;
Mat.-hes 89; Conservative 43; Mismatches 104; Indels 190; Gaps 23;
QY 5 FVTASGGQF-VLNLGPRYGGTNNYLSYOSHADVDDVLAKAQAAMLSVIRTWGTFIDIGS 63
Db 462 FLINSGVNEKVTHTPISEDYGNVFEYGMVQ-----NLSVF----- 496
QY 64 LQGSVPTIDGNKNGFYQWDPSTGAPAYND--GPTGLOGLDYAIASAAAHLGLRVIVLVT 121
Db 497 -----SGLDKGKNGYKKNIDP-----NDPNP----- 519
QY 122 NDWKEFGGDQYDKWYGLPYHDNFYTDPTQOAYKNWVHLLNRVNSITGVYKNDP--- 178
Db 520 -----GLPYR-HYYTDQSSQVPQN-----LNTPNPL-----YRNMPPQNS 552
QY 179 -----TIFAWELAN-----EPRCVGSGTL-----PTSGCTQATIVNNVDDQM 215
Db 553 HAIGNIIGCFMQANYNILSNVIVGACTRYDIYLLDKNGRTHVTSGFSPSATVLY----- 607
QY 216 SAYVKSIDPNHVMVSGDEGYI-----GSTQSGSWPYNDPSGDVNNALLRVK-----NID 266
Db 608 -----NPESIGLKVSAYATKCALPGDGLMRDPTVIYQRLRPAIGQNVFNVYD 658
QY 267 ECTYHL-----YPNYWGONADWGTQWIKDHIANAAGICKPTILEEFGHOTPD 313
Db 659 FNSKYFNVRGAAYFYQVINNFINSYQDTS-----KNGGGNATA-----KNMSGNLPE 705
QY 314 RDSVYQTWTVRTNGEAGNFWMLAGNVGQPPNYDG-FNV--YYPSSSTATVLA-SEA 369
Db 706 TINLYG-----YEVSQNVYR---YKNFLGTFESVARSWPTARGHLLADTYA 746
QY 370 LAITSG 375
Db 747 LAATTG 752

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 13, 2002, 10:49:26 ; Search time 13.829 Seconds
(without alignments)
1483.879 Million cell updates/sec

Title: US-09-917-378-4

Perfect score: 829

Sequence: 1 VSGGVKVOYKNNDSAPGDNQ.....TKVTYVYVNGRLVWCTEPSGT 154

Scoring table: BLOSUM62

Gapop 10.0 ; Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A_Geneseq_101002.*
1: /SID22/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2: /SID22/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3: /SID22/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4: /SID22/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5: /SID22/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6: /SID22/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7: /SID22/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8: /SID22/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9: /SID22/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10: /SID22/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11: /SID22/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12: /SID22/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13: /SID22/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14: /SID22/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15: /SID22/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16: /SID22/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17: /SID22/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18: /SID22/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19: /SID22/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20: /SID22/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21: /SID22/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22: /SID22/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
23: /SID22/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	374	45.1	551	AAW18790	Corrected Bacillus
2	373.5	45.1	700	AA13227	Novel endoglucanase
3	373	45.0	782	AA15625	Cellulase AP-1, A
4	369	44.5	1352	AAG63962	Amino acid sequenc
5	368	44.4	677	AAU98063	Bacillus subtilis
6	360	43.4	499	AA12122	NK-1 cellulase, B
7	357.5	43.1	476	AA154123	A mannanase-linker
8	357.5	43.1	493	AA128850	Pectate lyase-link
9	357.5	43.1	493	AA13218	Pectate lyase CBD
10	357	43.1	1350	22 AAG63963	Amino acid sequenc

11	356.5	43.0	167	17	AA95080	Cellulose binding
12	355.5	42.9	531	18	AAW15238	Scaffoldin protein
13	355.5	42.9	1853	19	AAW43108	C. thermocellum ce
14	349	42.1	616	20	AA13494	Truncated cellulase
15	349	42.1	616	23	AAE16325	Active cellulase h
16	348	42.0	1751	20	AA13493	Truncated cellulase p
17	348	42.0	1751	23	AAE16324	Active cellulase p
18	340	41.0	1426	20	AA13492	Truncated cellulase
19	340	41.0	1426	23	AAE16323	Active cellulase p
20	226	27.3	162	20	AAW90077	C. cellulovorans C
21	225	27.1	162	15	AAE63634	Cellulose binding
22	225	27.1	163	22	AAE05745	Clostridium cellu
23	225	27.1	328	22	AAE05749	Chimeric S peptide
24	225	27.1	341	22	AAE05747	Clostridium cellu
25	225	27.1	382	20	AA139552	Gaussia luciferase
26	225	27.1	382	23	AAE13383	Gaussia species CB
27	225	27.1	428	22	AAE05748	Clostridium cellu
28	224	27.0	156	20	AAW90080	C. cellulovorans C
29	224	27.0	190	22	AAE05746	Clostridium cellu
30	218	26.3	154	20	AAW90081	C. cellulovorans C
31	127	15.3	531	16	AAW1503	60 kD endoglucanase
32	127	15.3	532	12	AA13229	Endoglucanase enco
33	120.5	14.5	256	22	AAE81128	C17E2 OSPA constru
34	120.5	14.5	256	23	AAU97869	E. coli codon opti
35	101	12.2	646	22	AAE50866	Bacillus lichenifo
36	90.5	10.9	986	21	AA140440	Human brain-deri
37	90.5	10.9	1131	23	AAE23672	Human secretin rec
38	90.5	10.9	1346	22	AAU04567	Human G-protein co
39	90.5	10.9	1346	22	AAU04581	Human G-protein co
40	90.5	10.9	1371	22	AAU04570	Human G-protein co
41	85.5	10.3	665	15	AAE47189	Cellulase AEC3, A
42	85	10.3	2123	22	AAE00701	Moraxella catarrha
43	84.5	10.2	475	22	AAE92298	C glutamicum prote
44	84	10.1	434	19	AAW50008	Human hyaluronidas
45	84	10.1	1992	17	AAW04505	Moraxella 200 kDa

ALIGNMENTS

RESULT 1
AAW18790
ID AAW18790 standard; protein; 551 AA.
AC AAW18790;
XX
DT 18-NOV-1997 (first entry)
XX
DE Corrected Bacillus lautus (NCIMR 40250) endoglucanase Endo 3A.
XX
KW Endoglucanase; Endo 3A; formation: localised; variation:
KW colour density; surface; dye; fabric; family 5; cellulose;
KW hydrolysis; p-nitrophenyl-beta-1,4-cellobioside; stone wash;
KW blue jeans; back staining.
XX
OS Bacillus lautus.
XX
PN WO9709410-A1.
XX
PD 13-MAR-1997.
XX
PF 03-SEP-1996; 96WO-DK00364.
XX
PR 08-SEP-1995; 95DK-0000993.
XX
PA (NOVO) NOVO-NORDISK AS.
XX
PI Fich M, Onishi M, Schulein M, Toft AH;
XX WPI; 1997-192888/17.
XX
PT Localised variation of colour density in the surface of a dyed
cellulosic fabric - uses cellulase compsn. able to hydrolyse

PT p-nitrophenyl -beta-1,4-cellobioside
 XX Disclosure: Pages 15-17; 23pp; English.
 XX
 CC The present sequence is the corrected version of the incorrect
 CC Bacillus lautus (NCIMB 40250) endoglucanase Endo 3A described in
 CC WO9110732. Endo 3 can be used in novel method of forming localised
 CC colour density variation on the surface of a dyed cellulosic
 CC fabric. The method comprises agitating the fabric in an aqueous
 CC medium (pH 6.5 to 9.0) containing a family 5 cellulose,
 CC e.g. the present sequence, which can hydrolyse p-nitrophenyl-beta
 CC -1,4-cellobioside, or a family 7 cellulase, and a mechanical
 CC abrading agent or cellulose having abrading activity. Each
 CC cellulase displays 30 % or more of its maximum activity at pH 7.
 CC The process is useful to provide a stone washed look to blue jeans
 CC without back staining.
 XX
 XX Sequence 551 AA;
 SQ

Query Match 45.1%; Score 374; DB 18; Length 551;
 Best Local Similarity 47.1%; Pred. No. 7.2e-29;
 Matches 72; Conservative 33; Mismatches 42; Indels 6; Gaps 4;

Qy 2 SCGVKVOYKKNDSAPCDNQIKPGQLVNTGSSVDLSTVTYVYFTDRGGSTLVYNCW 61
 Db 402 TCNLVVOYKVGDTATDNOMKPSFNKIKNGTTPVNLGKLRYFTKD-GTADMSASF 460
 Qy 62 AAMCGNIRASFGSVNATPTADTYLQLSFT--GGTLAAGSGTGEIONRVKNSDWSNFDE 119
 Db 461 AQIGASVNSAAF--ANFTGSDTYDYVELSFGSISIPAGGTGDIQLRMKYKTDSNFNE 518
 Qy 120 TNDYSY-GTNTAFQDWTKVTVYVNGRLVMGTEP 151
 Db 519 ANDYSYDCAKTAYADNRVTLHQNGTLVWGTTTP 551

RESULT 2
 AAR13227
 ID AAR13227 standard; Protein: 700 AA.
 XX
 AC AAR13227;
 XX
 DT 14-OCT-1991 (first entry)
 XX
 DE Novel endoglucanase.
 XX
 KW Cellulase activity: detergent.
 XX
 OS Bacillus spp. NCIMB 40250.
 XX

Key Location/Qualifiers
 FT Peptide 1..31
 FT /label= signal sequence
 FT Protein 32..700
 FT /label= mature endoglucanase
 FT Cleavage-site 31..32
 XX
 PN WO9110732-A.
 XX
 XX
 PD 25-JUL-1991.
 XX
 XX
 PF 18-JAN-1991; 91WO-DK00013.
 XX
 XX
 PR 19-JAN-1990; 90DK-0000164.
 XX
 XX (NOVO) NOVO NORDISK A/S.
 PA
 XX Jorgensen PL, Schulein M, Hansen C;
 PI
 XX WPI: 1991-238020/32.
 DR
 DR N-PSDB; AAQ13001.
 XX
 PT Enzyme exhibiting cellulase activity from Bacillus sp. - is an

PT endo-glucanase, esp. useful for harshness redn. of cotton-contg.
 PT fabrics.
 XX
 PS Claim 1; Page 80; 96pp; English.
 XX
 CC The enzyme is produced by a strain of Bacillus spp. NCIMB 40250
 CC and exhibits an endoglucanase activity of at least 10 (pref. at
 CC least 25) carboxymethyl cellulose (CMC) endoase units per mg total
 CC protein under alkaline conditions. It is especially useful as a
 CC cellulolytic agent and has been found to be more stable during
 CC washing (60 mins. at 40 deg.) in the presence of conventional
 CC detergents than a commercial cellulase preparation. It may also
 CC show increased storage stability in liq. detergents contg.
 CC proteases. The sequence was deduced from the DNA (AAQ13001).
 CC See also AAR13228 and AAR13229.
 XX
 XX Sequence 700 AA;
 SQ

Query Match 45.1%; Score 373.5; DB 12; Length 700;
 Best Local Similarity 49.7%; Pred. No. 1.1e-28;
 Matches 77; Conservative 32; Mismatches 39; Indels 7; Gaps 5;

Qy 1 VSGGVKVOYKKNDSAPCDNQIKPGQLVNTGSSVDLSTVTYVYFTDRGGSTLVYNCW 59
 Db 549 VNSDLVVOYKDGDRNNATDNQIKPHFNQNKGTSPVDSLTLRYVFTKD-SSAAMNGW 607
 Qy 60 DWAAAGCGNIRASFGSVNATPTADTYLQLSFT--GGTLAAGSGTGEIONRVKNSDWSNF 117
 Db 608 DWAKGGSGNIQISFGNHNGA--DSDTYAEIAGFGSAGSIAEGGQSGEIQLRMSKADWSNF 665
 Qy 118 DETNDYSY-GTNTAFQDWTKVTVYVNGRLVMGTEP 151
 Db 666 NEANDYSFDGAKTAYIDMDRVTLYODGOLVWGIEP 700

RESULT 3
 AAR15625
 ID AAR15625 standard; Protein: 782 AA.
 XX
 AC AAR15625;
 XX
 DT 17-MAR-1992 (first entry)
 XX
 DE Cellulase AE-1.
 XX
 KW Detergents; pharmaceuticals; deinking; carboxymethylcellulose.
 XX
 OS Aeromonas strain no. 212.
 XX

JP03251174-A.
 XX
 PD 08-NOV-1991.
 XX
 PF 28-FEB-1990; 90JP-0045465.
 XX
 PR 28-FEB-1990; 90JP-0045465.
 XX
 XX (OJIP) OJI PAPER KK.
 XX
 XX WPI: 1991-373412/51.
 DR
 DR N-PSDB; AAQ15178.
 XX

Cellulase AE-1 for e.g. mfr. of pharmaceuticals and foodstuffs -
 of opt. pH when carboxymethylcellulose is used as substrate.
 PT
 XX
 XX Claim 2; Fig 3; 8pp; Japanese.
 XX

The sequence was deduced from the gene which was sequenced from
 CC plasmid, pAEC 1, prepd. by ligating chromosomal DNA contg. the
 CC gene (obtd. from Aeromonas) into pUC18. The protein has amol. wt.
 CC of 81,000 (SDS-PAGE) and an optimum pH near to 5 when carboxy-
 CC methylcellulose is the substrate. The N-terminal sequence: GHADT-
 CC has been confirmed by Edman degradation. The gene can be used to

CC produce recombinant enzyme which is used for the effective utilis-
CC action of biomass resources and the mfr. of pharmaceuticals and
CC foodstuffs, and also for the detergent and deinking of waste paper.
XX
SQ Sequence 782 AA:

Query Match 45.0%; Score 373; DB 12; Length 782;
Best Local Similarity 45.8%; Pred. No. 1.4e-28;
Matches 70; Conservative 30; Mismatches 49; Indels 4; Gaps 3;

QY 2 SCGVKVOYKNDSDAPCDNQIKPCLQLVNTGSSVDLSTVTVRWFTTRDGSSTLVNCDW 61
II : IIII : IIII : IIII : IIII : IIII : IIII : IIII : IIII : IIII :
Db 631 SCDLAVOYKTDGTDNAADNQKPHFNIVNKGAAVPLSELRLRYFTAD-GNDOLQYNCW 689

QY 62 AAMCCNIRASFGSVNPAATPTADTYLQLSP--TGGTLAAGSGTGEIONRVNKSDFNDE 119
I : II : III : IV : V : VI : VII : VIII : IX : X : XI : XII : XIII : XIV : XV :
Db 690 AMVCCSNLNCFAVKMNPKANADTYLEITFKAAGSLQAPAGKRGISOTRNHAGNWANLNE 749

QY 120 TNDYSYG-TNTAFODWTKVTYVYVNGRLVMGTEP 151
: IIII : II : III : IIII : IIII : IIII : IIII : IIII : IIII : IIII :
Db 750 SNDYSFDPTKTAVANNKRVTLYHNGTLVFCNEP 782

RESULT 4
AAG63962
ID AAG63962 standard; Protein: 1352 AA.
AC AAG63962;
XX
DT 29-OCT-2001 (first entry)
XX
DE Amino acid sequence of xyloglucanase enzyme.
XX
KW xyloglucanase; family 44; glycosyl hydrolase; detergent;
KW cellulosic fiber; textile scouring.
XX
OS Paenibacillus polymyxa.
XX
PN WO200162903-A1.
XX
PD 30-AUG-2001.
XX
PF 21-FEB-2001; 2001WO-DK00116.
XX
PR 24-FEB-2000; 2000DK-0000291.
XX
PA (NOVO) NOVOZYMES AS.
XX
PI Schnorr K, Jorgensen PL, Schuelein M;
XX
DR WPI; 2001-522819/57.
DR N-PSDB; AAH75059.
XX
PT New xyloglucanase enzyme belonging to glycosyl hydrolases family,
PT useful for detergent compositions, and textile or cellulose fiber
PT processing industries -
XX
PS Example 3; Page 81-85; 97pp; English.
XX
CC The present sequence represents a xyloglucanase of the invention. The
CC specification describes a xyloglucanase enzyme belonging to family 44
CC of glycosyl hydrolases and exhibiting a relative xyloglucanase activity
CC of at least 30% at pH 5-8. The enzyme exhibits high performance in
CC detergent compositions and prevents binding of certain soils to the
CC xyloglucan left in the cellulosic material. It is stable at pH 5-10 at
CC room temperature and has a half life of more than 50 days when incubated
CC in a full formulated liquid detergent at 30 plus degrees celcius. The
CC enzyme is used in detergent compositions, textile industry for improving
CC the properties of cellulosic fibers, yarn, and woven or non-woven
CC fabrics, preferably in textile scouring process, and in cellulose fiber
CC processing industry for ratting of fibers e.g. hemp, jute, flax, and
CC linen.
XX

SQ Sequence 1352 AA:

Query Match 44.5%; Score 369; DB 22; Length 1352;
Best Local Similarity 44.7%; Pred. No. 7.3e-28;
Matches 68; Conservative 33; Mismatches 47; Indels 4; Gaps 3;

QY 4 GVKVQYKNDSDAPCDNQIKPCLQLVNTGSSVDLSTVTVRWFTTRDGSSTLVNCDWAA 63
II : III : IIII : IIII : IIII : IIII : IIII : IIII : IIII : IIII :
Db 1200 GLLLQYRTADTKVNDNHLNPOFOIVNKGTTSPINELKIRYYTIDGDREQ-TFNCDYAT 1258

QY 64 MCGCNRASFGSVNPAATPTADTYLQLSFT--GGTLAAGSGTGEIONRVNKSDFNDETN 121
: III : IIII : IIII : IIII : IIII : IIII : IIII : IIII : IIII : IIII :
Db 1259 LSCSKLNGKLVKMEKATGADYVLEVSFNSDAGVLAPCGSGTDIQTIRHKTDMSYNESD 1318

QY 122 DYSY-GTNTAFODWTKVTYVYVNGRLVMGTEPS 152
IIII : IIII : IIII : IIII : IIII : IIII : IIII : IIII : IIII : IIII :
Db 1319 DYSYGTGTSFADHPKVTLYHNGVLVMGTEPT 1350

RESULT 5
AAU98063
ID AAU98063 standard; Protein: 677 AA.
AC AAU98063;
XX
DT 12-AUG-2002 (first entry)
XX
DE Bacillus subtilis spore coat protein cotG-CMCase fusion protein.
XX
KW Spore coat protein; spore surface; delivery system; bioconversion;
KW recombinant live vaccine; resistance property; protein microarray;
KW whole cell absorber; secretion signal; target signal; cell membrane;
KW surface display motif; cotG; CMCase; carboxymethyl cellulase.
XX
OS Chimeric - Bacillus subtilis.
OS Synthetic.
XX
PN WO200246388-A1.
XX
PD 13-JUN-2002.
XX
PF 07-DEC-2001; 2001WO-KR02124.
XX
PR 08-DEC-2000; 2000KR-0074835.
XX
PA (GENO-) GENOFOCUS CO LTD.
XX
PI Pan J, Choi S, Jung H;
XX
DR WPI; 2002-463830/49.
DR N-PSDB; ABK53202.
XX
PT Displaying a protein of interest on spore surface, by transforming a
PT host cell with a vector comprising a gene encoding a spore coat protein
PT and a gene encoding a protein of interest, useful for preparing protein
PT microarray -
XX
PS Example 4; Page 150; 158pp; English.
XX
CC The present invention relates to a new method for displaying a protein
CC of interest on a spore surface. The method of the invention involves
CC preparing a vector comprising a gene construct containing a gene
CC encoding a spore coat protein and a gene encoding a protein of interest,
CC transforming a host cell with the vector, displaying the protein of
CC interest on a surface of a spore of the host cell, and recovering the
CC spore. The method is useful for producing a delivery system for
CC recombinant live vaccine. The method is also useful for improving a
CC protein of interest using a resistance property of spore, in
CC bioconversion reactions, for preparing protein microarray, for
CC producing an antibody to antigen in vertebrates, and for preparing a
CC whole cell absorber. The spore coat proteins used circumvent a necessity
CC for passage across cell membrane, so that they do not need secretion
CC signal and target signal which are prerequisites of surface display

Qy 4 GVKVQYKKNDSAPGDNQIKPGLQLVNTGSSSVDLSTVTVRYWF-TRDGGSTLVYNCDDWA 62
1: |||| | |||: | : | : |||| || |||: | : | : |||: |
Db 353 GLSVOYKAGGGGVNSOIRPOLHIKNGGNATVDLKDVTARYWYNAKNGON--FQCTVA 408

CC and galactoglucmannans. The mannanase protein, or preparations
CC containing it, are used to improve properties of cellulosic or
CC synthetic fibres, yarn or (non)woven fabrics (removal of mannan-based
CC sizes or printing pastes). They are also used to degrade or modify
CC plant materials (particularly recycled waste paper, paper making pulps,
CC or material containing guar or locust bean gums (thickeners), or to
CC reduce viscosity of mannan-containing foods (or feeds). The mannanases
CC are also used to process coffee extracts (to inhibit gel formation);
CC in cleaning compositions (for machine washing of fabrics, as
CC hard-surface cleaners, for hand or machine dishwashing, also in oral,
CC dental, contact lens or body-care compositions) where they remove
CC mannan-containing soils and prevent binding of some soils to
CC cellulosics; and in fabric softeners. They can also be used in oil
CC well drilling to fracture subterranean formations.

XX Sequence 476 AA;
SQ
Query Match 43.1%; Score 357.5; DB 21; Length 476;
Best Local Similarity 42.3%; Pred. No. 2.7e-27;
Matches 69; Conservative 36; Mismatches 47; Indels 11; Gaps 3;

Qy 1 VSGGVKVOYKNNDSAPGDNQIKPGLQLVNTGSSVDLSTVTVRVWFTTRDGSSTLVYVNC 60
Db 314 VSGNLKVEFYNSPDDTNSINPOKVTNTGSSAIDLSKLTLYRYVTVDGOKDQTFW-CD 372
Qy 61 WAAM-----CCGNIRASFGSVNPATPTADTYLQLSFTGCTLAAGSGTGEIONRVNK 111
Db 373 HAALIGSNGSYNGITSNVKGTFVKMSSTNNADTYLEISFTGCTLEPCAHV-QIOGFEAK 431
Qy 112 SDWSNFDNTDYSYGTWATFQDTWKTVYVYNGRLVWGTPEPSGT 154
Db 432 NDWSNYTQSDNYSFKRSQFVWDQVAYLNGVLVWGKEPGGS 474

RESULT 8
AA28850
ID AAY28850 standard; Protein: 493 AA.
XX
AC AAY28850;
XX
DT 17-JAN-2000 (first entry)
XX
DE Pectate lyase-linker-CBD fusion protein.
XX
KW Pectate lyase-linker-CBD fusion protein; ATCC 14580; plasmid pMB914;
KW cellulose binding domain; CBD; linker; transformed; detergent;
KW cellulosic fibre; yarn; degradation; recycled waste paper; animal feed;
KW paper-making pulp; retting process; processing; wine; juice.

XX Chimeric - Bacillus licheniformis.
OS Chimeric - Clostridium thermocellum.
XX WO9927083-A1.
XX
PD 03-JUN-1999.
XX
PF 24-NOV-1998; 98WO-DK00514.
XX
PR 24-NOV-1997; 97DK-0001344.
PR 06-MAY-1998; 98US-0073684.
XX
PA (NOVO) NOVO-NORDISK AS.
XX
PI Andersen LN, Bjornvad ME, Lange NEK, Schnorr K, Schuelein M;
XX WPI: 1999-610578/52.
DR N-PSDB; AAX90978.
XX
XX New isolated pectate lyase enzymes -
PT Example 4; Page 85-86; 93pp; English.
PS
XX The present sequence is a Pectate lyase-linker-CBD fusion protein.

CC Plasmid pMB914 was constructed using pectate lyase gene from
CC Bacillus licheniformis, ATCC 14580 and cellulose binding domain (CBD)
CC gene from Clostridium thermocellum YS fused with a linker molecule.
CC Bacillus subtilis was transformed with plasmid pMB914 for expression of
CC the fusion protein. Pectate lyase can be used in detergent compositions,
CC for cleaning hard surfaces, for machine treatment of fabrics, for
CC improving the properties of cellulosic fibres, yarn, woven or non-woven
CC fabric, for the degradation of plant material e.g. recycled waste paper,
CC mechanical paper-making pulps or fibres subjected to retting process, for
CC preparing animal feed and for processing wine or juice.

XX Sequence 493 AA;
SQ
Query Match 43.1%; Score 357.5; DB 20; Length 493;
Best Local Similarity 42.3%; Pred. No. 2.8e-27;
Matches 69; Conservative 36; Mismatches 47; Indels 11; Gaps 3;

Qy 1 VSGGVKVOYKNNDSAPGDNQIKPGLQLVNTGSSVDLSTVTVRVWFTTRDGSSTLVYVNC 60
Db 331 VSGNLKVEFYNSPDDTNSINPOKVTNTGSSAIDLSKLTLYRYVTVDGOKDQTFW-CD 389
Qy 61 WAAM-----CCGNIRASFGSVNPATPTADTYLQLSFTGCTLAAGSGTGEIONRVNK 111
Db 390 HAALIGSNGSYNGITSNVKGTFVKMSSTNNADTYLEISFTGCTLEPCAHV-QIOGFEAK 448
Qy 112 SDWSNFDNTDYSYGTWATFQDTWKTVYVYNGRLVWGTPEPSGT 154
Db 449 NDWSNYTQSDNYSFKRSQFVWDQVAYLNGVLVWGKEPGGS 491

RESULT 9
AA43218
ID AAY43218 standard; Protein: 493 AA.
XX
AC AAY43218;
XX
DT 13-JAN-2000 (first entry)
XX
DE Pectate lyase CBD fusion protein sequence.
XX
KW Pectate lyase; polysaccharide lyase; enzyme; pectin degradation;
KW polygalacturonide; detergent composition; hard surface treatment;
KW cellulosic fibre; plant material degradation; recycled waste paper;
KW mechanical paper-making pulp; wine processing; cellulose binding domain;
KW ClpB.
XX
OS Clostridium thermocellum.
OS Bacillus sp.
OS Synthetic.
XX
PN WO9927084-A1.
XX
PD 03-JUN-1999.
XX
PF 24-NOV-1998; 98WO-DK00515.
XX
PR 24-NOV-1997; 97DK-0001343.
PR 24-NOV-1997; 97DK-0001344.
PR 06-MAY-1998; 98US-0073684.
PR 02-NOV-1998; 98US-0184217.
XX
PA (NOVO) NOVO-NORDISK AS.
XX
PI Andersen LN, Schuelein M, Lange NEK, Bjornvad ME, Moller S;
XX Glad SOS, Kauppinen MS, Schnorr K, Kongsbaek L;
XX WPI: 1999-610579/52.
DR N-PSDB; AAZ31562.
XX
XX New isolated pectate lyase enzymes -
PT Claim 26; Page 106-108; 113pp; English.
PS
XX

This sequence is a fusion protein comprising a Bacillus species pectate lyase of the invention, fused via a linker to the Clostridium thermocellum C1pB protein cellulose binding domain (CBD). The pectate lyase enzymes are obtained from Bacillus agaradhaerens (BA), Bacillus licheniformis (BL), Bacillus halodurans (BH) and other Bacillus species. The pectate lyase enzymes can be used for degrading pectin, pectate and polygalacturonides. They can be used in detergent compositions, for cleaning hard surfaces, for machine treatment of fabrics, for improving the properties of cellulosic fibres, yarn, woven or non-woven fabric, for the degradation of plant material (e.g. recycled waste paper, mechanical paper-making pulps or fibres subjected to a retting process, for preparing animal feed or for processing wine or juice. DNA encoding the enzymes can also be used for the production of transgenic plants.

Query Match 43.1%; Score 357.5; DB 20; Length 493;
Best Local Similarity 42.3%; Pred. No. 2.8e-27;
Matches 69; Conservative 36; Mismatches 47; Indels 11; Gaps 3;

QY 1 VSGVKVQYKKNDSAPGDNOIKPGLQLVNTGSSVDLSTVTYRYWFTRDGSSSTLVNCD 60
DB 331 VSGNLKVEYNPSDTHNSINPQKVTNTGSSAIDLSKLTLYRYTYVDGKDQTFW-CD 389
QY 61 WAAM-----GCGNIRASFGSVNPATPTADTYLQLSFTGGTAAAGSGTGEIQNRVVK 111
DB 390 HAAITGSGSYNGCITSNVKATFVKSSSTNNADTYLEISFTGTLPEGAHV-QIOGRFAK 448
QY 112 SOWSNFDEINDSYCNTAFQDWTKVTYVNGRLVWGTEPSGT 154
DB 449 NDWSNTQSDNYSFKRSQSFVEMDQVATYLVNGRLVWGTEPGGS 491

RESULT 10
AAG63963
ID AAG63963 standard; Protein; 1350 AA.

XX AAG63963;
XX 29-OCT-2001 (first entry)
XX Amino acid sequence of xyloglucanase enzyme.
XX xyloglucanase; family 44; glycosyl hydrolase; detergent;
KW cellulosic fiber; textile scouring.

XX Paenibacillus polymyxa.

XX Key Location/Qualifiers
FH Misc-difference 1347 /note= "Glu encoded by GAACCG"

XX WO200162903-A1.
XX 30-AUG-2001.
XX 21-FEB-2001; 2001WO-DK00116.
XX 24-FEB-2000; 2000DK-0000291.

XX (NOVO) NOVOZYMES AS.
XX Schnorr K, Jorgensen PL, Schuelein M;
XX WPI; 2001-522819/57.
DR N-PSDB; AAH75060.

XX New xyloglucanase enzyme belonging to glycosyl hydrolases family,
PT useful for detergent compositions, and textile or cellulose fiber
PT processing industries -

XX Disclosure; Page 86-89; 97pp; English.

CC The present sequence represents a xyloglucanase of the invention. The
CC specification describes a xyloglucanase enzyme belonging to family 44
CC of glycosyl hydrolases and exhibiting a relative xyloglucanase activity
CC of at least 30% at pH 5-8. The enzyme exhibits high performance in
CC detergent compositions and prevents binding of certain soils to the
CC xyloglucan left in the cellulosic material. It is stable at pH 5-10 at
CC room temperature and has a half life of more than 50 days when incubated
CC in a full formulated liquid detergent at 30 plus degrees Celsius. The
CC enzyme is used in detergent compositions, textile industry for improving
CC the properties of cellulosic fibers, yarn, and woven or non-woven
CC fabrics, preferably in textile scouring process, and in cellulose fiber
CC processing industry for retting of fibers e.g. hemp, jute, flax, and
XX linen.

XX Sequence 1350 AA;

Query Match 43.1%; Score 357; DB 22; Length 1350;
Best Local Similarity 44.7%; Pred. No. 1.2e-26;
Matches 67; Conservative 32; Mismatches 47; Indels 4; Gaps 3;

QY 4 GYKVOYKKNDSAPGDNOIKPGLQLVNTGSSVDLSTVTYRYWFTRDGSSSTLVNCDMAA 63
DB 1199 GLVLQTRTADTNVNDHLNPHFOILNKGITISVPINELKIRYYITIDGDRQ-TFNCUYAV 1257
QY 64 MCGNIRASFGSVNPATPTADTYLQLSFT--GCTLAAGSGTGEIQNRVKNKSDNSFDET 121
DB 1258 LSCSKLNGKLVKMDKAATGADYYLEVFSNSDAGVLSGSGTGCTRIHKADSNYNESD 1317
QY 122 DYSY-CTNTAFQDWTKVTYVNGRLVWGTE 150
DB 1318 DYSYKGTQTSFDDHTKATLYHNGVLVWGTE 1347

RESULT 11
AAR95080
ID AAR95080 standard; peptide; 167 AA.

XX AAR95080;
XX 30-OCT-1996 (first entry)
XX Cellulose binding domain.

XX Cellulose binding domain; CBD; haptens; moiety; biotin; avidin;
KW streptavidin; affinity chromatography; cell separation;
KW cell immobilisation; protein immobilisation; enzyme immobilisation;
KW multienzyme reactors; signal immunoassays; drug delivery; pesticide;
KW cellulose; chitin.

XX Clostridium thermocellum.

XX Key Location/Qualifiers
FH Misc-difference 139 /note= "Unidentified amino acid."

XX WO9613524-A1.
XX 09-MAY-1996.
XX 26-OCT-1995; 95WO-US13813.
XX 27-OCT-1994; 94IL-O111415.

XX (TECR) TECHNION RES & DEV FOUND LTD.
PA (UYRA-) UNIV RAMOT APPLIED RES & IND DEV LTD.
PA (YEDA) YEDA RES & DEV CO LTD.

XX Bayer EA, Lamed R, Morag E, Shoham Y, Wilchek M;
XX WPI; 1996-239453/24.

XX Modified cellulose-binding domain mols. - having attached haptens,
PT partic. biotin, to provide an affinity system for, e.g. sepsins.,

PT assays, reactors, delivery etc.
XX Claim 12; Page 34-35; 53pp; English.
PS A modified cellulose binding domain (CBD) or fraction of it, linked
CC to a hapten moiety via one or more Cys or Lys residues and complexes
CC comprising the biotinylated CBD and a biotin-binding molecule
CC selected from modified or unmodified avidin or streptavidin or an
CC anti-biotin antibody, can be used in affinity chromatography, cell
CC separation, cell, protein and enzyme immobilisation, selective
CC removal of biological materials, multi-enzyme reactors, signal
CC immunoassays and drug (e.g. pesticide) delivery. The CBD molecule
CC can be modified with the hapten without affecting its high affinity
CC for cellulose, or chitin.
XX
SQ Sequence 167 AA:

Query Match 43.0%; Score 356.5; DB 17; Length 167;
Best Local Similarity 42.3%; Pred. No. 8.6e-28;
Matches 69; Conservative 36; Mismatches 47; Indels 11; Gaps 3;

QY 1 VSGGVKVOYKKNDSAPGDNQIKPGLQLVNTGSSVDLSTVTYVRYWFTRDGGSSTLVYVNC 60
DB 5 VSGNLKVEFYNSNPSTTNSINPQKVTNTGSSAIDLSKLTLYYYTVVVGQKDOTFW-CD 63
QY 61 MAAM-----GCCNIRASFGSVNPATPTADTYLQLSFTGCTLAAGSGTGEIQNRVVK 111
DB 64 HAAIIGSNGSYNGITSNVKCTFVKMSSSTNNADTYLEISFTGCTLEPGAHV-QIOGRFAK 122
QY 112 SDWSNFDENDSYGTNTAFQDWTKVTVYVNGRLVNGTEPSGT 154
DB 123 NDWSNTQSDNYSFKXSOFVEMDQVATYLVNGVLVNGKEPGGS 165

RESULT 12
AAW15238
ID AAW15238 standard; protein; 531 AA.
XX
AC AAW15238:
XX
DT 28-JAN-1998 (first entry)
DE Scaffoldin protein from Clostridium thermocellum.
XX
KW Dockerin; CelB; CipA; scaffoldin; cellulose binding domain;
KW chromatographic separation; soluble substrate modification; CBD;
KW multi-enzyme delivery system; animal feed; paper production;
KW plant protection; pest control.
XX
OS Clostridium thermocellum.
XX
FH Key Location/Qualifiers
FT Region 1..153
FT /label= internal_repeat_element_1
FT Region 154..306
FT /label= internal_repeat_element_2
FT Domain 239..531
FT /label= cellulose_binding
FT /note= "only 60% of the CBD is present"
XX
DN W09714789-A2.
XX
XX
PD 24-APR-1997.
XX
XX 16-OCT-1996; 96WO-US16485.
XX
XX 17-NOV-1995; 95US-0559968.
XX
XX 17-OCT-1995; 95US-0005701.
XX
XX (GENV) GENECOR INT INC.
XX
XX Bott RR, Clarkson KA, Fowler T, Liu C, Ward M, Xia H;
XX

DR WPI: 1997-245106/22.
XX
PT Composition with enzymes non-covalently bound to a peptide backbone
PT - used as a multi-enzyme delivery system, e.g. in food processing,
PT textiles and pest control
XX
PS Claim 10; Fig 6; 29pp; English.
XX
CC This protein sequence represents a scaffoldin based on the CipA protein
CC of Clostridium thermocellum. The scaffoldin protein is used in a novel
CC composition that comprises at least 2 enzymes non-covalently bound to a
CC peptide backbone (i.e. present sequence). The scaffoldin comprises a
CC number of internal repeating units and at least one cellulose binding
CC domain (CBD). The CBD may be altered to modify its affinity for
CC cellulose, which may be desirable where cellulose binding would be
CC disadvantageous. The enzyme is bound to the scaffoldin by a dockerin
CC region of the enzyme, which binds to a repeating element of the
CC scaffoldin. The dockerin is preferably Cels (AAW15237) or CelB
CC (AAW15236). The composition can be used in reducing allergenicity,
CC producing synergistic effects, and facilitating selective modification
CC of substrate. By taking advantage of the cellulose binding domain of the
CC complex, the complex could be immobilised for use in chromatographic
CC separations or for soluble substrate modification. By adding the
CC scaffoldin domain, it is possible to recover enzymes, or to quantify
CC the amount of an enzyme in a solution. The composition could also be
CC used in a multi-enzyme delivery system which could be used in the food
CC industry, in food processing, animal feed, textiles, bioconversion,
CC pulp and paper production, plant protection and pest control, as a wood
CC preservative, topical lotions, and biomass conversions.
XX
SQ Sequence 531 AA:

Query Match 42.9%; Score 355.5; DB 18; Length 531;
Best Local Similarity 42.3%; Pred. No. 4.9e-27;
Matches 69; Conservative 36; Mismatches 47; Indels 11; Gaps 3;

QY 1 VSGGVKVOYKKNDSAPGDNQIKPGLQLVNTGSSVDLSTVTYVRYWFTRDGGSSTLVYVNC 60
DB 312 VSGNLKVEFYNSNPSTTNSINPQKVTNTGSSAIDLSKLTLYYYTVVVGQKDOTFW-CD 370
QY 61 MAAM-----GCCNIRASFGSVNPATPTADTYLQLSFTGCTLAAGSGTGEIQNRVVK 111
DB 371 HAAIIGSNGSYNGITSNVKCTFVKMSSSTNNADTYLEISFTGCTLEPGAHV-QIOGRFAK 429
QY 112 SDWSNFDENDSYGTNTAFQDWTKVTVYVNGRLVNGTEPSGT 154
DB 430 NDWSNTQSDNYSFKXSOFVEMDQVATYLVNGVLVNGKEPGGS 472

RESULT 13
AAW43108
ID AAW43108 standard; Protein; 1853 AA.
XX
AC AAW43108:
XX
DT 16-OCT-1998 (first entry)
DE C. thermocellum cellulosome integrating protein.
XX
KW Multimer; enzyme; complex; protein-protein interaction; dockerin domain;
KW cohesin domain; catalytic subunit; scaffold subunit; SdbA; synergistic;
KW cellulosome integrating protein; scaffoldin dockerin binding protein.
XX
OS Clostridium thermocellum.
XX
PN FR2748479-Al.
XX
PD 14-NOV-1997.
XX
XX 10-MAY-1996; 96FR-0005854.
XX
XX 10-MAY-1996; 96FR-0005854.
XX


```

PN   US6294366-B1.
XX
PD   25-SEP-2001.
XX
PF   19-AUG-1998;    98US-0136574.
XX
PR   19-SEP-1997;    97US-0932571.
XX
PA   (CLRN ) CLARIANT FINANCE BVI LTD.
XX
PI   Farrington GK, Anderson P, Bergquist P, Daniels R, Gibbs MD;
PT   Morgan H, Williams DP;
XX
DR   WPI; 2002-081780/11.
XX
N-PSDB; AAD26568.
XX
XX New cellulase active protein,'useful in textile processing or
PT commercial detergents, e.g. for improving the feel or appearance of
PT cotton-containing fabrics,'is stable under conditions of alkaline pH
PT and elevated temperatures'.
XX
XX Disclosure: Column 75-78; 61pp; English.
XX
CC The present invention relates to a cellulase active protein, which is
CC substantially free of proteinases of native thermophilic and
CC alkaliphilic origin, where the cellulase active protein consists of the
CC CelB5 amino acid sequence. The cellulase active protein is useful for
CC treating cellulosic materials including cotton-containing fabrics, as
CC detergent additives. The cellulase active protein is also useful for
CC improving the feel and/or appearance of cotton-containing fabrics, for
CC removing surface fibers from cotton-containing knits or for imparting
CC stone-washed appearance to cotton-containing denims. The present proteins
CC are stable under condition of alkaline pH and elevated temperatures,
CC thus suitable for textile processing and in commercial detergents.
XX The present sequence is E3/B5 hybrid protein.
XX
SQ Sequence 616 AA;
Query Match 42.1%; Score 349; DB 23; Length 616;
Best Local Similarity 43.8%; Pred. No. 2.6e-26;
Matches 71; Conservative 26; Mismatches 49; Indels 16; Gaps 4;
QY 1 VSGCGVKVOYKKNDSAPGCDNOIKPGLOLVNTGSSSDVLDSTVTVRVRYWFTRDGSGSSTLYVNC D 60
Db 1 MGSGGVKVLKNNETYSATGSIRPFWKIVNGSGSSVDLSRVKIRVYWTVYDGDQPQSAV-CD 59
      ||| :|||::|::| |::| |::|::| |::|::| |::|::| |::|::| |::|::|
QY 61 WAAMGCCGNIRASFGSNVPATPATDYTLQLSLTGGT--LAAGSGTGEICNRVNKNSDWSNFD 118
Db 60 WAOIGASNVTFFNKLLSSGVSCADYYLEVGFSGGAGOLQGDKTDGIQVRFNKNDSWNFIN 119
      || :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||
QY 119 ETNDYS-----YGTNTAFODWTKVTVVYNGLRWGMTEPSG 153
Db 120 QADDWSMLQSMTNYGEN-----AKVTLYVDGVLVWGPEPG 155

```

Search completed: November 13, 2002, 11:51:40
Job time : 14.829 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 13, 2002, 11:50:41 ; Search time 7.19109 Seconds
(without alignments)
2058.756 Million cell updates/sec

Title: US-09-917-378-4
Perfect score: 829
Sequence: 1 VSGGVKVOYKNDSAPGDNQ.....TKVTYVYNGRLVMGTPEPSGT 154

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR73:*
1: pir1.*
2: pir2.*
3: pir3.*
4: pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	394	47.5	145	2 A41897	cellulase homolog
2	373.5	45.1	700	2 B41897	cellulase (EC 3.2.)
3	370	44.6	508	2 G69593	cellulase (EC 3.2.)
4	368	44.4	499	2 A26111	cellulase (EC 3.2.)
5	365	44.0	508	2 A26874	cellulase (EC 3.2.)
6	360	43.4	499	2 A27198	cellulase (EC 3.2.)
7	359.5	43.4	1711	2 T31337	1,4-beta-glucanase
8	355.5	42.9	1854	2 S36859	c1pA protein - Clo
9	352	42.5	1331	2 A48954	mannan endo-1,4-be
10	352	42.5	1742	2 T17120	cellulase (EC 3.2.)
11	347	41.9	1039	2 S02711	cellulase (EC 3.2.)
12	344	41.5	1779	2 T31085	cellulase (EC 3.2.)
13	343.5	41.4	504	2 S34744	xyylanase - Caldice
14	341	41.1	915	2 A43802	cellulase (EC 3.2.)
15	335	40.4	505	2 S39962	endoglucanase - Er
16	309	37.3	486	2 I40548	bifunctional cellu
17	299.5	36.1	1483	2 C97012	probably celluloso
18	290	35.0	1162	2 T30433	scaffolding protei
19	287.5	34.7	1230	2 PC4746	cellulose 1,4-beta
20	286	34.5	586	2 PC6006	scaffolding protei
21	285	34.4	986	2 S12021	thermoactive cellu
22	257	31.0	879	2 A47704	endoglucanase I (E
23	225	27.1	1848	2 A44140	cellulose-binding
24	136.5	16.5	1045	2 A39199	endoglucanase B (E
25	101.5	12.2	636	2 JC5874	cellulase (EC 3.2.)
26	97.5	11.8	269	2 H95976	endo-beta-1,3-1,4-
27	95.5	11.5	712	2 F97012	hypothetical prote
28	92.5	11.2	1428	2 AC2224	hypothetical prote
29	90.5	10.9	618	2 T08605	hypothetical prote

30	90	10.9	268	2 S34804	exok protein - Mhi
31	89.5	10.8	725	2 JC1300	endo-beta-1,4-gluc
32	87.5	10.6	2348	2 AD1841	hypothetical prote
33	86	10.4	820	2 T14879	hypothetical prote
34	85.5	10.3	829	2 AG0869	probable outer mem
35	85	10.3	582	2 F70675	hypothetical prote
36	82.5	10.0	319	2 T43040	hypothetical prote
37	82.5	10.0	411	2 T21758	large repetitive p
38	82	9.9	3283	2 AC1018	hypothetical prote
39	81.5	9.8	547	2 T25478	hypothetical prote
40	81.5	9.8	582	1 HNNZMM	hemagglutinin-neur
41	81.5	9.8	582	1 HNNZSB	hemagglutinin-neur
42	81.5	9.8	698	2 D90771	hypothetical prote
43	81.5	9.8	698	2 H85633	hypothetical prote
44	81	9.8	958	2 A82583	conserved hypothet
45	80	9.7	506	2 F83369	hypothetical prote

ALIGNMENTS

RESULT 1

A41897
cellulase homolog - Bacillus lautus (fragment)
C:Species: Bacillus lautus
C:Date: 17-Feb-1994 #sequence_revision 17-Feb-1994 #text_change 15-Oct-1999
C:Accession: A41897; S27498
R:Hansen, C.K.; Diderichsen, B.; Jorgensen, P.L.
J. Bacteriol. 174, 3522-3531, 1992
A:Title: cels from Bacillus lautus PL236 encodes a novel cellulose-binding endo-beta-
A:Reference number: A41897; MUID:92276330; PMID:1592807
A:Accession: A41897
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-145 <HAN>
A:Cross-references: EMBL:M76588; NID:g142661; PIDN:AAA22302.1; PID:g142662
A:Experimental source: PL236
A>Note: sequence extracted from NCBI backbone (NCBIP:104604)

Query Match 47.5%; Score 394; DB 2; Length 145;
Best Local Similarity 50.7%; Pred. No. 1.7e-28;
Matches 75; Conservative 31; Mismatches 36; Indels 6; Gaps 4;

QY	7	VOYKNDSAPGDNQIKPGQLQVNTGSSVDLSVTYVYVWFTRDGSGSTLVYVNCDAAMGC	66
DB	1	LOYRAADTNAADNQIKPSFNKNNGTSVDLSLTKIRYFTKDGSAVWVW-IDMAOLCG	59
QY	67	GNIRASFGSVNPATPTADTYLQLSFT--GGTLAAGSGTGEIONRVNKSQMSNFDETNDYS	124
DB	60	SNIOISFG--NHTGTNSDTYVELSFSSSEAGSIAAGSGGSETOLRMSKTDMSNPNENDYS	117
QY	125	Y-GTNTAFODWTKVYVYVNGRLVMGTPE	151
DB	118	FDGKTAFADMDRVVLYQNGQIVMGCTAP	145

RESULT 2

B41897
cellulase (EC 3.2.1.4) - Bacillus lautus
N:Alternate names: endo-1,4-beta-glucanase
C:Species: Bacillus lautus
C:Date: 10-Mar-1994 #sequence_revision 10-Mar-1994 #text_change 15-Oct-1999
C:Accession: B41897; S27499
R:Hansen, C.K.; Diderichsen, B.; Jorgensen, P.L.
J. Bacteriol. 174, 3522-3531, 1992
A:Title: cels from Bacillus lautus PL236 encodes a novel cellulose-binding endo-beta-
A:Reference number: A41897; MUID:92276330; PMID:1592807
A:Accession: B41897
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-700 <HAN>
A:Cross-references: EMBL:M76588; NID:g142661; PIDN:AAA22303.1; PID:g142663
A:Experimental source: PL236

[illegible][illegible]


```

A:Reference number: A27198; MUID:87190397; PMID:3106035
A:Accession: A27198
A:Molecule type: DNA
A:Residues: 1-499 <NAK>
A:Cross-references: GB:M28332; NID:q142670; PIDN:AAA22307.1; PID:q142671
A:Experimental source: strain IF03034
C:Function:
A:Description: hydrolysis of 1,4-beta-D-glucosidic linkages in beta-D-glucan
A:Pathway: cellulose degradation
C:Keywords: extracellular protein; glycosidase; hydrolase; polysaccharide
F:1-36/Domain: signal sequence #status predicted <SIC>

Query Match 43.4%; Score 360; DB 2; Length 499;
Best Local Similarity 43.6%; Pred. No. 8.6e-25;
Matches 65; Conservative 32; Mismatches 48; Indels 4; Gaps

Qy 4 GVKVQYKNDSPACDQNIKPGIQLVNTGSSVDLSLTVTYRYWF-TRDGGSTFLVYVNCDA 62
Db 353 GISVQYKAGDGVNSQIRPOLHIKKNGNATVLDKDVARTARYWYNAKNKGN---FDCDTA 409
Qy 63 AMGCGNIRASFGSVNPATPYTADTYLQLSFTGCTLAAGSGTGEIQNRVAKSOMSNFDETN 122
Db 410 QIGCGNLTHKFVTLHKPKGADTYLELGFKCTLTSPGASTGNIQLRLHNDWMSNYAQSGI 469
Qy 123 YSYCTNTAFQDWTKVTVYVNGRLVWGTEP 151
Db 470 YSFQSNFTFKTKITLYHOGKLWGTET 498

RESULT 7
T31337
1,4-beta-glucanase (EC 3.2.1.-) - Anaerocellum thermophilum (fragmentL)
C:Species: Anaerocellum thermophilum
C:Date: 02-Sep-2000 #sequence_revision 02-Sep-2000 #text_change 02-Sep-2000
C:Accession: T31337
A:Molecule type: DNA
A:Residues: 1-1711 <ZVE>
A:Cross-references: EMBL:286105; NID:e1071329; PID:e350354; PIDN:CAB06786.1
C:Genetics:
A:Gene: cea
C:Keywords: glycosidase; hydrolase

Query Match 43.4%; Score 359.5; DB 2; Length 1711;
Best Local Similarity 46.6%; Pred. No. 4e-24;
Matches 76; Conservative 25; Mismatches 45; Indels 17; Gaps

Qy 1 VSGG-VKVOYKNNDSAPGDNQIKPQLQLVNTGSSVDLSLTVTYRYWFTRDGGSTFLVYVNC 59
Db 882 VAGGQIKVLVYANKENTSTNTIRPMLKVVNTGSSSIDLSRVTIRVYVTVGDGKAQSAIS- 940
Qy 60 DWAMCGGNIRASFGSVNPATPYTADTYLQLSFTGCT--LAAGSGTGEIQNRVAKSOMSNF 117
Db 941 DWAIQIGASNVTFKFKVLKSSVSVDYVLEIGFKSGAGQLQAGKDTGEIQIRFNKSDWSNY 1000
Qy 118 DETNDYS-----YCTNTAFQDWTKVTVYVNGRLVWGTEPSPG 153
Db 1001 NOGNDWSWQSMNTNYGEN-----VKVTAYIDGVLVWQGEPSG 1037

RESULT 8
S36859
c1pA protein - Clostridium thermocellum
N:Alternate names: probable cellulosome protein large chain SL
C:Species: Clostridium thermocellum
C:Date: 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 01-Dec-2000
C:Accession: S36859; S33527; S28659; T18261

```

R:Gorngross, U.T.; Demain, A.L.
submitted to the EMBL Data Library, January 1993
A:Reference number: S36859
A:Accession: S36859
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1854 <GER>
A:Cross-references: EMBL:L08665
R:Gorngross, U.T.; Romaniec, M.P.M.; Kobayashi, T.; Huskisson, N.S.; Demain, A.L.
Mol. Microbiol. 8, 325-334, 1993
A:Title: Sequencing of a Clostridium thermocellum gene (cipA) encoding the cellulosomal
A:Reference number: S33527; MUID:93302508; PMID:8316083
A:Accession: S33527
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1615,1617-1854 <GE2>
A:Cross-references: EMBL:L08665
R:Fujino, T.; Beguin, P.; Aubert, J.P.
FEMS Microbiol. Lett. 94, 165-170, 1992
A:Title: Cloning of a Clostridium thermocellum DNA fragment encoding polypeptides that b
A:Reference number: S25767
A:Accession: S25767
A:Molecule type: DNA
A:Residues: 1216-1232, 'Y', 1234-1241, 'T', 1243-1319, 'R', 1321-1615,1617-1854 <FUJ>
A:Cross-references: EMBL:X67406
R:Bequin, P.
submitted to the EMBL Data Library, August 1992
A:Reference number: S28659
A:Accession: S28659
A:Molecule type: DNA
A:Residues: 1216-1232, 'Y', 1234-1241, 'T', 1243-1319, 'R', 1321-1615,1617-1768, 'R', 1770-1854
A:Cross-references: EMBL:X67406
R:Fujino, T.; Beguin, P.; Aubert, J.P.
J. Bacteriol. 175, 1891-1899, 1993
A:Title: Organization of a Clostridium thermocellum gene cluster encoding the celluloso
e.
A:Reference number: 218847; MUID:93209931; PMID:8458832
A:Accession: T18261
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1821-1854 <FU2>
A:Cross-references: EMBL:X67506; NID:9296879; PID:9296880; PIDN:CAA47840.1
C:Genetics:
A:Gene: cipA
Query Match 42.9%; Score 355.5; DB 2; Length 1854;
Best Local Similarity 42.3%; Pred. No. 1e-23;
Matches 69; Conservative 36; Mismatches 47; Indels 11; Gaps 3;
Qy 1 VSGGVKVOYKNNDSAPGDNQIKPGLQLVNTGSSVDLSTVTYVRYWFTTRDGGSTLVYKCD 60
Db 365 VSGNLKVEFYNSNPSTTNSINPQFKVTNTGSSAIDLKSLTURYTYTVDGQKDQTFW-CD 423
Qy 61 WAAM-----GCCNIRASFGSVNPATPTADTYLQLSFTGGTLAAGSGTGEIQNRVVK 111
Db 424 HAAIIGSNGSYNGITSNVKGTFVRKMSSTNNADTYLEISFTGGTLEPGAHV-OIQGRFAK 482
Qy 112 SDWSNFDETNDYSYGTNTAFQDWTKTVTYVNGRLVWGTEPSGT 154
Db 483 NDWSNYTOSNDYSFKSASQFVEMDQVTAYLNGVLVWGKEPGGS 525
RESULT 9
A48954
mannan endo-1,4-beta-mannosidase (EC 3.2.1.78) - Caldocellum saccharolyticum
N:Alternate names: beta-mannanase
C:Species: Caldocellum saccharolyticum
C:Date: 19-Dec-1993 #sequence_revision 18-Nov-1994 #text_change 08-Oct-1999
C:Accession: A48954; B43745
R:Gibbs, M.D.; Saul, D.J.; Luthi, E.; Bergquist, P.L.
Appl. Environ. Microbiol. 58, 3864-3867, 1992
A:Title: The beta-mannanase from "Caldocellum saccharolyticum" is part of a multidomain
A:Reference number: A48954; MUID:93119139; PMID:1476429

A:Accession: A48954
A:Status: preliminary
A:Molecule type: nucleic acid
A:Residues: 1-1331 <GIB>
A:Cross-references: GB:L01257; NID:g144290; PIDN:AAA71887.1; PID:g144291
A:Note: sequence extracted from NCHI backbone (NCBIN:121576, NCBI:P:121577)
R:Luthi, E.; Bhana Jasmal, N.; Grayling, R.A.; Love, D.R.; Bergquist, P.L.
Appl. Environ. Microbiol. 57, 694-700, 1991
A:Title: Cloning, sequence analysis, and expression in Escherichia coli of a gene cod
A:Reference number: A43745; MUID:91247819; PMID:2039230
A:Accession: B43745
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-337, 'PPRQHQHRQ' <LUE>
A:Cross-references: EMBL:M36063; NID:g144292; PIDN:AAA72861.1; PID:g144294
A:Note: the authors translated the codon CAC for residue 262 as Glu
A:Note: this sequence has been revised in reference A48954
C:Keywords: glycosidase; hydrolase; polysaccharide degradation
Query Match 42.5%; Score 352; DB 2; Length 1331;
Best Local Similarity 45.8%; Pred. No. 1.4e-23;
Matches 71; Conservative 30; Mismatches 50; Indels 4; Gaps 3;
Qy 3 GGVKVOYKNNDSAPGDNQIKPGLQLVNTGSSVDLSTVTYVRYWFTTRDGGSTLVYKCDNA 62
Db 568 GOIKVLYANKETNSTTNTIRPWLKVNVSGSSIDLSRVTIRYWTVDGERAQSAIS-DWA 626
Qy 63 AMGCCNIRASFGSVNPATPTADTYLQLSFTGCT--LAAGSGTGEIQNRVKNKSDWSNFDET 120
Db 627 QIGASNVTFKFKLVSSVSGADYILEIGPKSGAGLOPKDTGCTOIRFNKSDWSNFOG 686
Qy 121 NDYSY-GTNTAFQDWTKTVTYVNGRLVWGTEPSGT 154
Db 687 NDWSIQSNTSYGENEKVTAYIDGVLVWGQEPSGT 721
RESULT 10
T17120
cellulase (EC 3.2.1.-) precursor, thermoactive - Caldocellum saccharolyticum
C:Species: Caldocellum saccharolyticum
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 17-Nov-2000
C:Accession: T17120; A43745
R:Te'o, V.S.; Saul, D.J.; Bergquist, P.L.
Appl. Microbiol. Biotechnol. 43, 291-296, 1995
A:Title: Cella, another gene coding for a multidomain cellulase from the extreme therm
A:Reference number: 218698; MUID:95336703; PMID:7612247
A:Accession: T17120
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1742 <TEO>
A:Cross-references: EMBL:L32742; NID:g537499; PID:g537500; PIDN:AAA91086.1
R:Luthi, E.; Bhana Jasmal, N.; Grayling, R.A.; Love, D.R.; Bergquist, P.L.
Appl. Environ. Microbiol. 57, 694-700, 1991
A:Title: Cloning, sequence analysis, and expression in Escherichia coli of a gene cod
A:Reference number: A43745; MUID:91247819; PMID:2039230
A:Accession: A43745
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1516-1544, 'A', 1546-1742 <LUE>
A:Cross-references: EMBL:M36063; NID:g144292; PIDN:AAA72860.1; PID:g144293
C:Genetics:
A:Gene: cellA
C:Keywords: glycosidase; hydrolase
Query Match 42.5%; Score 352; DB 2; Length 1742;
Best Local Similarity 45.5%; Pred. No. 2e-23;
Matches 71; Conservative 30; Mismatches 51; Indels 4; Gaps 3;
Qy 2 SGGVKVOYKNNDSAPGDNQIKPGLQLVNTGSSVDLSTVTYVRYWFTTRDGGSTLVYKCDW 61
Db 704 SQIKVLYANKETNSTTNTIRPWLKVNVSGSSIDLSRVTIRYWTVDGERAQSAIS-DW 762
Qy 62 AAMGCCNTRASFSGSVNPATPTADTYLQLSFTGCT--LAAGSGTGEIQNRVKNKSDWSNFD 119

Db 763 A Q I G A S N V T F K F V K L S S V S G A D Y L E I G F K S G A C O L Q P G K D T G E I Q I R F N K D W S N Y N Q 822

Qy 120 T N D Y S Y - G T N T A F Q D M T K V T V Y V N G R L V M G T E P S C T 154

Db 823 G N D W S I Q S M T S Y G E N E K V T A Y I D G V L V M G Q E P S C T 858

RESULT 11

\$O2711

cellulase (EC 3.2.1.4) precursor - Caldocellum saccharolyticum

N:Alternate names: endo-1,4-beta-glucanase

N:Contains: cellulase (EC 3.2.1.4); cellulose 1,4-beta-cellobiosidase (EC 3

C:Species: Caldocellum saccharolyticum

C:Date: 07-Sep-1990 #sequence_revision 07-Sep-1990 #text_change 08-Oct-1999

C:Accession: S02711

R:Saul, D.J.; Williams, L.C.; Love, D.R.; Chamley, L.W.; Bergquist, P.L.

Nucleic Acids Res. 17, 439, 1989

A:Title: Nucleotide sequence of a gene from Caldocellum saccharolyticum enc

A:Reference number: S02711; MUID:85098398; PMID:2789517

A:Accession: S02711

A:Molecule type: DNA

A:Residues: 1-1039 <SAU>

A:Cross-references: EMBL:X13602; NID:g40645; PIDN:CAA31936.1; PID:g40646

C:Genetics:

A:Gene: cellB

C:Function:

A:Description: hydrolysis of 1,4-beta-D-glucosidic linkages in beta-D-gluca

A:Pathway: cellulose degradation

C:Superfamily: Streptomyces endo-1,4-beta-xylanase A homology

C:Keywords: glycosidase; hydrolase; polysaccharide degradation

F:1-28/Domain: signal sequence #status predicted <SIG>

F:29-1039/Product: cellulase #status predicted <MAT>

F:72-373/Domain: Streptomyces endo-1,4-beta-xylanase A homology <SXY>

Query Match 41.9%; Score 347; DB 2: Length 1039;

Best Local Similarity 45.5%; Pred. No. 3e-23;

Matches 70; Conservative 30; Mismatches 50; Indels 4; Gaps

Qy 3 GGVKVQYKNNDSPAGNDQIKPGLQIVNTGSSVDLSLTVTRYWFTROGSGSTLVYNCDA 62

Db 420 GQIKVLYANKETNSTNTIRPLKVVNSGSSSIDLSRVTIRYVWTVDCERAAQSASV-DWA 478

Qy 63 AMCCGNIRASFSGVNPATPTADTYLQLSFTGCT---LAAGSGTGEIQNRVKNKSDWSNFEET 120

Db 479 QIGASNVTFKPVKLSVSSVGADYYLEIGFKSGAGLQPKGTGTEIRFNKSDWSNYSNQ 538

Qy 121 NDXYSGTNTAFQDMTKVTVVYVNGRLVMGTEPSG 153

Db 539 NDWSLQISMTSYGENEKVTAYIDGVLVWGQEPG 572

RESULT 12

T31085

xylanase - Caldicellulosiruptor sp.

C:Species: Caldicellulosiruptor sp.

C:Date: 02-Sep-2000 #sequence_revision 02-Sep-2000 #text_change 02-Sep-2000

C:Accession: T31085

R:Morris, D.D.; Gibbs, M.D.; Ford, M.; Thomas, J.; Bergquist, P.L.

submitted to the EMBL Data Library, December 1997

A:Description: Family 10 and 11 xylanase genes from Caldicellulosiruptor sp

A:Reference number: 220972

A:Accession: T31085

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-1779 <MOR>

A:Cross-references: EMBL:AF036924; NID:g2760905; PID:g2760909; PIDN:AAB95322

C:Genetics:

A:Note: xynC

Query Match 41.5%; Score 344; DB 2: Length 1779;

Best Local Similarity 44.0%; Pred. No. 1.1e-22;

Matches 70; Conservative 25; Mismatches 48; Indels 16; Gaps

Search completed: November 13, 2002, 11:57:35
Job time : 9.19109 secs

GenCore version 5.1.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 13, 2002, 11:47:41 : Search time 3.54023 Seconds
(without alignments)
1804.220 Million cell updates/sec

Title: US-09-917-378-4

Perfect score: 829

Sequence: 1 VSGGVKQYKNDSDAPGDNQ.....TKTVVYVNGRLVWGTEPSCGT 154

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 20000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_40.*

pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	394	47.5	145	1 YCEA_PAEIA	P29718 paenibacill
2	373.5	45.1	700	1 GUN2_BACSU	P29719 paenibacill
3	370	44.6	499	1 GUN2_BACSU	P10475 bacillus su
4	368	44.4	499	1 GUN3_BACSU	P23549 bacillus su
5	365	44.0	499	1 GUN1_BACSU	P07983 bacillus su
6	357.5	43.1	772	1 CIPB_CLOTM	O01866 clostridium
7	355.5	42.9	1853	1 CIPB_CLOTM	O06851 clostridium
8	352	42.5	1331	1 MAND_CALSA	P22533 caldocellum
9	352	42.5	1742	1 GUN2_CALSA	P22534 caldocellum
10	347	41.9	1039	1 GUN2_CALSA	P10474 c endogluc
11	343.5	41.4	504	1 GUN2_ERWCA	O59395 erwinia car
12	335	40.4	505	1 GUNV_ERWCA	O47096 erwinia car
13	295.5	35.6	914	1 GUN2_CLOSR	P50900 clostridium
14	285	34.4	986	1 GUN2_CLOSR	P23659 clostridium
15	257	31.0	879	1 GUN1_CLOTM	O02934 clostridium
16	225	27.1	1848	1 CBPA_CLOCL	P38058 clostridium
17	149.5	18.0	444	1 GUNN_ERWCA	O59394 erwinia car
18	136.5	16.5	1045	1 GUNB_CELFI	P26225 cellulomona
19	101.5	12.2	636	1 GUN4_BACSU	P28622 bacillus sp
20	99	11.9	880	1 GUN4_THEFU	P26221 thermomonas
21	97.5	11.8	269	1 EXOK_RHIME	P33693 rhizobium m
22	89.5	10.8	725	1 GUNK_CLOCE	P37700 clostridium
23	81.5	9.8	582	1 HEMA_MUMPM	P11235 mumps virus
24	79.5	9.6	698	1 YMCA_ECOLI	P75882 escherichia
25	79.5	9.6	1200	1 ICEN_PSESY	P06820 pseudomonas
26	79	9.5	238	1 GUB_PAEPO	P45797 paenibacill
27	78.5	9.5	1210	1 ICEN_PSEFL	P09815 pseudomonas
28	77.5	9.3	498	1 VNUC_IAUSS	P18073 influenza a
29	77.5	9.3	582	1 HEMA_MUMPR	P10866 mumps virus
30	77.5	9.3	1567	1 ICEN_XANCT	P18127 xanthomonas
31	77	9.3	237	1 GUB_PAEPA	P23904 paenibacill
32	77	9.3	282	1 POR2_XENLA	P81004 xenopus lae
33	77	9.3	1006	1 BGAL_ASPNG	P29853 aspergillus

34	76	9.2	247	1 FLAI_THEVO	P57719 thermoplasm
35	76	9.2	1157	1 C9CA_BACTO	O45733 bacillus th
36	75.5	9.1	644	1 XYND_CELFI	P54865 cellulomona
37	74.5	9.0	488	1 PHB_ALCPA	P12625 alicigenes
38	74.5	9.0	498	1 VNUC_IADRA	P18069 influenza a
39	74.5	9.0	498	1 VNUC_IAPOW	P18071 influenza a
40	74.5	9.0	498	1 VNUC_IAME3	O07545 influenza a
41	74.5	9.0	582	1 HEMA_MUMPI	P19762 mumps virus
42	74	8.9	1148	1 ICEK_PSESX	O30611 pseudomonas
43	74	8.9	1244	1 SLAI_YEAST	P32790 saccharomyc
44	74	8.9	3462	1 RELN_RAT	P58751 rattus norv
45	73.5	8.9	242	1 GUB_BACSU	P04957 bacillus su

ALIGNMENTS

RESULT 1

ID	YCEA_PAEIA	STANDARD;	PNT;	145 AA.
AC	P29718;			
DT	01-APR-1993 (Rel. 25, Created)			
DT	01-APR-1993 (Rel. 25, Last sequence update)			
DT	16-OCT-2001 (Rel. 40, Last annotation update)			
DE	Hypothetical protein in CELA 5' region (Fragment).			
OS	Paenibacillus lautus (Bacillus lautus).			
OC	Bacteria; Firmicutes; Bacillales; Paenibacillaceae; Paenibacillus.			
OX	NCBI_TaxID=1401;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=PL236;			
RX	MEDLINE=92276330; PubMed=1592807;			
RA	Hansen C.K., Joergensen P.L., Diderichsen B.;			
RT	"celA from Bacillus lautus PL236 encodes a novel cellulose-binding			
RL	J. Bacteriol. 174:3522-3531(1992).			

CC	-----
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration
CC	between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC	the European Bioinformatics Institute. There are no restrictions on its
CC	use by non-profit institutions as long as its content is in no way
CC	modified and this statement is not removed. Usage by and for commercial
CC	entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC	or send an email to license@isb-sib.ch).
CC	-----
DR	EMBL: M76588; AAA23202.1; -
DR	PIR: A41897; A41897.
DR	PIR: S27498; S27498.
DR	HSSP: O06851; INBC.
DR	InterPro: IPR001956; CBD_3.
DR	Pfam: PF00942; CBM_3; 1.
DR	ProDom: PD001947; CBD_3; 1.
KW	Hypothetical protein.
FT	NON_TER 1
SQ	SEQUENCE 145 AA; 15782 MW; 9514E3A71B10AEB CRC64;

Query Match 47.5%; Score 394; DB 1; Length 145;
Best Local Similarity 50.7%; Pred. No. 5.8e-30;
Matches 75; Conservative 31; Mismatches 36; Indels 6; Gaps 4;

QY	7	VQYKNDSDAPDNOIKPCLQLVNTGSSSSVDLSTVTYVWFTRDGGSSFLVYNCDAAMGC	66
DB	1	LQYRAADTNAADNOIKPSPFKNNKNGTSVLDLTKIRYFTKDGSAAVNGW-IDMAOLGG	59
QY	67	GNIRASEGVSNPATPTADTYLQLSFT--GGTLAGGSGTGTQNRVKNKSDMSNFDETNDYS	124
DB	60	SNIOISFG--NHTGNTSDTYVELSFSSEAGSIAAGSGSGETQLRMSKTRDSNPFNEANDYS	117
QY	125	Y-GTNTAFQDWTQVTVVYVNGRLVWGTEP	151
DB	118	FDGKTAFADMDRVLLYQNGQIVNGTAP	145

```

RESULT 2
GUNA_PAE1A
ID GUNA_PAE1A STANDARD: PRT: 700 AA.
AC P29719;
DT 01-APR-1993 (Rel. 25, Created)
DT 01-APR-1993 (Rel. 25, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE Endoglucanase A precursor (EC 3.2.1.4) (Endo-1,4-beta-glucanase)
DE (Cellulase A) (EG-A).
GN CELA.
OS Paenibacillus lautus (Bacillus lautus).
OC Bacteria; Firmicutes; Bacillales; Paenibacillaceae; Paenibacillus.
OX NCBI_TaxID=1401;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=PL236;
RX MEDLINE=92276330; PubMed=1592807;
RA Hansen C.K., Diderichsen B., Joergensen P.L.;
RT "celA from Bacillus lautus PL236 encodes a novel cellulose-binding
RT endo-beta-1,4-glucanase.";
RL J. Bacteriol. 174:3522-3531(1992).
CC -!- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic
CC linkages in cellulose.
CC -!- PTM: A SHORT FORM (EGA-S) ARISES FROM POSTTRANSLATIONAL
CC PROTEOLYSIS OF APPROXIMATELY 150 AA AT THE C-TERMINUS OF EGA-L.
CC -!- SIMILARITY: BELONGS TO CELLULASE FAMILY J (FAMILY 44 OF GLYCOSYL
CC HYDROLASES).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M76588; AAA22303.1; -.
DR PIR: B41897; B41897.
DR PIR: S27499; S27499.
DR HSP: Q06851; INBC.
DR InterPro: IPR001956; CBD_3.
DR Pfam: PF00942; CBM_3; 1.
DR ProDom: PD001947; CBD_3; 1.
KW Cellulose degradation; Hydrolase; Glycosidase; Signal.
FT SIGNAL.
FT CHAIN 1 33
FT DOMAIN 34 700 ENDOGLUCANASE A.
FT ACT_SITE 213 213 BY SIMILARITY.
SQ SEQUENCE 700 AA; 76910 MW; 3D5C8CADA53EEE0F CRC64;

Query Match 45.1%; Score 373.5; DB 1; Length 700;
Best Local Similarity 49.7%; Pred. No. 3; le-27;
Matches 77; Conservative 32; Mismatches 39; Indels 7; Gaps 5;

QY 1 VSGGVKVOYKND-SAPGDNQIKPGLQLVNTGSSVDLSVTVTYVWFTRDGGSSSTLVYNC 59
| : ||||| : | : ||||| : | : ||||| : | : ||||| : | : ||||| : | :
Db 549 VNSDLVVQYKCDRRNATNDQIKPHFNIONKCTSPVDSLSLTRYFTKD-SSAANGWI 607
| : ||||| : | : ||||| : | : ||||| : | : ||||| : | : ||||| : | :
QY 60 DWAMCGNIRASFGSNVPATPATYTLQLSFT--GGTLAAGSGTGEIQNRVYKSDWSNF 117
| : ||||| : | : ||||| : | : ||||| : | : ||||| : | : ||||| : | :
Db 608 DWAKLGGSNIQISFGNHGA--DSDTYAELGFSGAGSIAEGGQSGEIQLRMSKADWSNF 665
| : ||||| : | : ||||| : | : ||||| : | : ||||| : | : ||||| : | :
QY 118 DETNDYSY-GTNTAFQDWTKVTYVYNGRLVWGTEP 151
| : ||||| : | : ||||| : | : ||||| : | : ||||| : | : ||||| : | :
Db 666 NEANDYSFGAKTAYIDMRVTLYODGOLVWGIEP 700

RESULT 3
GUN2_BACSU
ID GUN2_BACSU STANDARD: PRT: 499 AA.
AC P10475;
DT 01-JUL-1989 (Rel. 11, Created)
DT 01-JUL-1989 (Rel. 11, Last sequence update)

```

```

DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Endoglucanase precursor (EC 3.2.1.4) (Endo-1,4-beta-glucanase)
DE (Carboxymethyl-cellulase) (CMCAGE) (Cellulase).
GN BGLC OR GLD OR EGLS.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=1423;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=PAP115;
RX MEDLINE=87066783; PubMed=3024130;
RA Mackay R.M., Lo A., Willick G., Zuker M., Baird S., Dove M.,
RA Moraneilli F., Seligy V.;
RT "Structure of a Bacillus subtilis endo-beta-1,4-glucanase gene.";
RL Nucleic Acids Res. 14:9159-9170(1986).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=CK-2;
RX MEDLINE=95225656; PubMed=7710280;
RA Lindahl V., Aa K., Tronsmo A.;
RT "Nucleotide sequence of an endo-beta-1,4-glucanase gene from Bacillus
RT subtilis CK-2.";
RL Antonie Van Leeuwenhoek 66:327-332(1994).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RX MEDLINE=97124194; PubMed=8969507;
RA Rose M., Entian K.D.;
RT "New genes in the 170 degree region of the Bacillus subtilis genome
RT encode DNA gyrase subunits, a thioredoxin, a xylanase and an amino
RT acid transporter.";
RL Microbiology 142:3097-3101(1996).
RN [4]
RP SEQUENCE OF 30-45.
RC STRAIN=CK-2;
RX MEDLINE=95225655; PubMed=7710279;
RA Aa K., Flensburg R., Lindahl V., Tronsmo A.;
RT "Characterization of production and enzyme properties of an
RT endo-beta-1,4-glucanase from Bacillus subtilis CK-2 isolated from
RT compost soil.";
RL Antonie Van Leeuwenhoek 66:319-326(1994).
CC -!- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic
CC linkages in cellulose.
CC -!- SIMILARITY: BELONGS TO CELLULASE FAMILY A (FAMILY 5 OF GLYCOSYL
CC HYDROLASES).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: Z29076; CAA82317.1; -.
DR EMBL: X04869; CAA28392.1; -.
DR EMBL: X67044; CAA47429.1; -.
DR EMBL: Z73234; CAA97610.1; ALT_INIT.
DR EMBL: Z99113; CAB13696.1; ALT_INIT.
DR PIR: A26114; A26114.
DR HSP: O85465; IAH3.
DR Subtilist; BG10437; bg1C.
DR InterPro: IPR001956; CBD_3.
DR InterPro: IPR001547; GH_5.
DR Pfam: PF00150; cellulase; 1.
DR Pfam: PF00942; CBM_3; 1.
DR ProDom: PD001947; CBD_3; 1.
DR PROSITE: PS00659; GLYCOSYL-HYDROL_F5; 1.
KW Cellulose degradation; Hydrolase; Glycosidase; Signal;
KW Complete proteome.
FT SIGNAL 1 29
FT CHAIN 30 499
FT ACT_SITE 169 169 ENDOGLUCANASE.
FT PROTON DONOR (BY SIMILARITY).

```

[illegible]

```

RESULT 4
GUN3_BACSU          STANDARD;          PRT;    499 AA.
ID   GUN3_BACSU
AC   P23549;
DT   01-NOV-1991 (Rel. 20, 'Created)
DT   01-NOV-1991 (Rel. 20, Last sequence update)
DT   01-FEB-1995 (Rel. 31, Last annotation update)
DE   Endoglucanase precursor (EC 3.2.1.4) (Endo-1,4-beta-glucanase)
DE   (Carboxymethyl-cellulase) (CMCase) (Cellulase).
DE   BGCC.
OS   Bacillus subtilis.
OC   Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX   NCBI_TaxID=1423;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BSE616;
RX   MEDLINE=91299280; PubMed=1368694;
RA   Park S.H., Kim H.K., Pack M.Y.;
RT   "Characterization and structure of the cellulase gene of Bacillus
RT   subtilis BSE616."
RL   Agric. Biol. Chem. 55:441-448(1991).
CC   -1- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic
CC   linkages in cellulose.
CC   -1- SIMILARITY: BELONGS TO CELLULASE FAMILY A (FAMILY 5 OF GLYCOSYL
CC   HYDROLASES).
CC   -----
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/
CC   or send an email to license@isb-sib.ch).

```

CC	EMBL; D01057; BAA00859.1; -	
DR	PIR; JN0111; JN0111.	
DR	HSP; O85465; 1A3H.	
DR	InterPro; IPR001956; CBD_3.	
DR	InterPro; IPR001547; GH_5.	
DR	Pfam; PF00150; cellulase; 1.	
DR	Pfam; PF00942; CBM_3; 1.	
DR	ProDom; PD001947; CBD_3; 1.	
DR	PROSITE; PS00659; GLYCOSYL_HYDROL_F5; 1.	
KW	Cellulose degradation; Hydrolase; Glycosidase; Signal.	
	SIGNAL	1 29
	FT	
	CHAIN	30 499
FT	ACT_SITE	169 169
FT	ACT_SITE	257 257
FT	DOMAIN	350 499
FT	SEQUENCE	499 AA: 55169 MW: 2E821E3D8BBACA04 CRC64;
CC		
DR	ENDOGALACTANASE.	
DR	PROTON DONOR (BY SIMILARITY).	
DR	NUCLEOPHILE (BY SIMILARITY).	
DR	CELLULOSE-BINDING (BY SIMILARITY).	

	Query Match	44.4%	Score 368;	DB 1:	Length 499;
	Best Local Similarity	43.6%;	Pred. No.	6.7e-27;	
	Matches	65;	Conservative	31;	Mismatches 51; Indels 2; Gaps 1;
Qy	4	GKVOYKNNDSSAPCDNOIKPCQLQLVNTGSSVDLSTVTVRYWFTDCGGSSLTVYNCDHAA	63	:	: :
Dd	353	GISVOYRAGDGSMSNQIRPOLQIKNGNTVVLDKDVTARYWNAKNKGNQV--DCDYAQ	410	:	: :
Qy	64	MCCGNIASFGSNVPATPTADTYLQLSFTGTGAAGSGTGEIQNRVNKSMDSEFENDY	123	:	: :
Dd	411	LGCQNVTYKFVTLHKPKOGADTYLELGFKNCTLPAGASTGCIQLRLHNDDHSNYAQS	470	:	: :
Qy	124	SYGTNPATPDQTKVTYYVINGRLVMWGTEPS	152	:	: :
Dd	471	SFFKSNFTFKTTKITLYDOCKLIWGTEPN	499	:	: :

```

RESULT 5
GUNI_BACSU          STANDARD;          PRT;    499 AA.
ID    GUNI_BACSU
AC    P07983;
DT    01-AUG-1988 (Rel. 08, Created)
DT    01-NOV-1991 (Rel. 20, Last sequence update)
DT    01-OCT-1994 (Rel. 30, Last annotation update)
DE    Endoglucanase precursor (EC 3.2.1.4) (Endo-1,4-beta-glucanase)
DE    (Cellulase).
DE    BGLC OR GLD.
OS    Bacillus subtilis.
OC    Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX    NCBI_TaxID=1423;
RN    [1]
RP    SEQUENCE FROM N.A.
RC    STRAIN=DLG;
RX    MEDLINE=87194581; PubMed3106328;
RA    Robson L.M., Chambliss G.H.;
RT    "Endo-beta-1,4-glucanase gene of Bacillus subtilis DLG.";
RL    J. Bacteriol. 169:2017-2025(1987).
CC    -I- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic
CC        linkages in cellulose.
CC    -I- SIMILARITY: BELONGS TO CELLULASE FAMILY A (FAMILY 5 OF GLYCOSYL
CC        HYDROLASES).

```

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL; M16185; AAA32496.1; ALT_INIT.
 PIR; A26874; A26874.
 HSSP; O85465; IA3H.
 InterPro; IPR001956; CBD_3.
 InterPro; IPR001547; GH_5.
 Pfam; PF00150; cellulase; 1.
 Pfam; PF00942; CBW_3; 1.
 ProDom; PD001947; CBD_3; 1.
 PROSITE; PS00659; GLYCOSYL_HYDROL_F5; 1.
 Cellulose degradation; Hydrolase; Glycosidase; Signal.
 SIGNAL 1 29
 CHAIN 30 499
 ACT_SITE 169 169
 ACT_SITE 257 257
 DOMAIN 350 499
 SEQUENCE 499 AA; 55187 MW; 339D04EE95A63BE1 CRC64;

Query Match	44.0%	Score 365;	DB 1;	Length 499;
Best Local Similarity	44.7%	Pred. No. 1.3e-26;		
Matches 67;	Conservative 31;	Mismatches 48;	Indels 4;	Gaps 2;
QY	4	GVKVOYKNNDSPGCDNQIIRPGLOLVNFGSSVDLSTVTVRYWYF-TRDGGSSLTLYVNCDMA	62	
			:	: : : : :

```

Db 353 GVSQYKAGDGRVNSQIRPOLHIKKNNGNATVDLKQVYRWYVNVKNKGN----FDCDYA 409
QY 63 AMCCGNIRASFGSVNPATPTADTYLQLSFTGGTAAAGSGTGETONRVNKSQWSNFDDETND 122
    IIIII: : : : IIIII: : III: : III: : : IIIII: : :
Db 410 QMCCGNLTHKFTVYLLHKKPGKADTYLELGFKGTGLSPGASIGNIQLKLNHDDWSNYYAQSGD 469
QY 123 YSYGNTAFQDWTKVTVVYVNGRLVWCTEPS 152
    II: : : : III: : : IIIII:
Db 470 YSFFQSNTFKTKKTLTYHQKLLJMCETPN 499

RESULT 6
CIPB_CLOTM
ID CIPB_CLOTM STANDARD: PRT: 772 AA.
AC Q01866;
DT 01-OCT-1994 (Rel. 30, Created)
DT 01-OCT-1994 (Rel. 30, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Cellulosomal scaffolding protein B (Cellulosomal glycoprotein sl/sL)
DE (Cellulose integrating protein B) (Fragment).
GN CIPB.
OS Clostridium thermocellum.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1515;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93146373; PubMed=1490597;
RA Poole D.M., Morag E., Lamed R., Bayer E.A., Hazlewood G.P.,
  Gilbert H.J.;
RT "Identification of the cellulose-binding domain of the cellulosome
  subunit Sl from Clostridium thermocellum YS."
RL FEMS Microbiol. Lett. 78:181-186(1992).
CC -1- FUNCTION: ACTS AS A SCAFFOLDING PROTEIN IN THE CELLULOSE. IT
  PROMOTES BINDING OF CELLULOSE TO THE CATALYTIC DOMAINS OF THE
  CELLULOYTIC ENZYMES PROBABLY THROUGH THE BINDING OF THE NINE
  REPEATED DOMAINS WITH THE DUPLICATED SEGMENT OF 24 AMINO ACIDS
  PRESENT IN CATALYTIC SUBUNITS OF THE CELLULOSE.
CC -1- SUBCELLULAR LOCATION: CELL SURFACE.
CC -1- SIMILARITY: A 24 RESIDUES DOMAIN IS REPEATED TWICE IN THIS ENZYME.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
  between the Swiss Institute of Bioinformatics and the EMBL outstation -
  the European Bioinformatics Institute. There are no restrictions on its
  use by non-profit institutions as long as its content is in no way
  modified and this statement is not removed. Usage by and for commercial
  entities requires a license agreement (See http://www.isb-sib.ch/announce/
  or send an email to license@isb-sib.ch).
CC
CC -----
DR EMBL; X68233; CAA48312.1; -
DR HSP; Q06851; INBC.
DR InterPro; IPR001956; CBD_3.
DR InterPro; IPR002102; Cohesin.
DR InterPro; IPR002105; Dockerin_1.
DR InterPro; IPR002048; EF-hand.
DR Pfam; PF00404; Dockerin_1; 2.
DR Pfam; PF00942; CBM_3; 1.
DR Pfam; PF00963; Cohesin; 3.
DR ProDom; PD001947; CBD_3; 1.
DR PROSITE; PS00018; EF_HAND; UNKNOWN_1.
DR PROSITE; PS00448; CLOS_CELLULOSE_RPT; 2.
KW Cellulose degradation; Cell wall; Glycoprotein; Repeat.
FT NON_TER 1 1
FT DOMAIN 1 607 3 x 147 AA APPROXIMATE REPEATS.
FT REPEAT <1 80 1 (PARTIAL).
FT REPEAT 94 240 2.
FT REPEAT 462 607 3..
FT DOMAIN 81 93 PRO/THR-RICH.
FT DOMAIN 241 272 PRO/THR-RICH.
FT DOMAIN 273 439 CELLULOSE-BINDING.
FT DOMAIN 440 461 PRO/THR-RICH.
FT DOMAIN 710 766 2 x 24 AA APPROXIMATE REPEATS.

```

```

FT REPEAT 710 733 2-1.
FT REPEAT 743 766 2-2.
SQ SEQUENCE 772 AA; 82491 MM; BBF06DE5E094FE10 CRC64;

Query Match 43.1%; Score 357.5; DB 1; Length 772;
Best Local Similarity 42.3%; Pred. NO. 1.le-25;
Matches 69; Conservative 36; Mismatches 47; Indels 11; Gaps 3;

QY 1 VSGGVKVOYKNDSAPCDNOIKPCLQLVNTGSSVDLSLTVTYVYWFTRDGGSSFLVYVNC 60
    IIII: : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 277 VSNLKVFEYNSNPSTTNSINPOFKVNTGSSAIDLSKLTLYRYTYVDGKQDTFW-CD 335
    IIII: : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 61 WAAM-----CGGNIRASFGSVNPATPTADTYLQLSFTGGTAAAGSGTGETONRVNK 111
    II: : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 336 HAAIIGSNQSYNGITSNVCKTFVKMSSSTNNADTYLEISFTGCTLEPGAHV-OIQGRFAK 394
    III: : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 112 SDWSNDEINQSYGNTAFQDWTKVTVVYVNGRLVWCTEPSGT 154
    IIII: : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 395 NQWSNTQSDNYSFKRSQFVENDQVTAYLNGVLVWKEPGGS 437

RESULT 7
CIPB_CLOTM
ID CIPB_CLOTM STANDARD: PRT: 1853 AA.
AC Q06851;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Cellulosomal scaffolding protein A precursor (Cellulosomal
  glycoprotein sl/sL) (Cellulose integrating protein A) (Cohesin).
GN CIPA.
OS Clostridium thermocellum.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1515;
RN [1]
RP SEQUENCE FROM N.A.; AND SEQUENCE OF 29-40 AND 42-43.
RX STRAIN-ATCC 27405 / DSM 1237;
RX MEDLINE=93302508; PubMed=8316083;
RA Gerngross U.T., Romaniec M.P.M., Kobayashi T., Huskisson N.S.,
  Demain A.L.;
RT "Sequencing of a Clostridium thermocellum gene (cipa) encoding the
  cellulosomal SL-protein reveals an unusual degree of internal
  homology."
RL Mol. Microbiol. 8:325-334(1993).
RN [2]
RP SEQUENCE OF 1820-1853 FROM N.A.
RX MEDLINE=93209931; PubMed=8458832;
RA Fujino T., Beguin P., Aubert J.-P.;
RT "Organization of a Clostridium thermocellum gene cluster encoding the
  cellulosomal scaffolding protein Cipa and a protein possibly involved
  in attachment of the cellulosome to the cell surface."
RL J. Bacteriol. 175:1891-1899(1993).
RN [3]
RP X-RAY CRYSTALLOGRAPHY (2.15 ANGSTROMS) OF 184-321.
RX MEDLINE=97238934; PubMed=9083107;
RA Shimon L.J.W., Bayer E.A., Morag E., Lamed R., Sharon Y.,
  Frolow F.;
RT "A cohesin domain from Clostridium thermocellum: the crystal
  structure provides new insights into cellulosome assembly."
RL Structure 5:381-390(1997).
RN [4]
RP X-RAY CRYSTALLOGRAPHY (1.8 ANGSTROMS) OF 368-522.
RX MEDLINE=97076134; PubMed=8918451;
RA Torno J., Lamed R., Chirino A.J., Morag E., Sharon Y.,
  Steitz T.A.;
RT "Crystal structure of a bacterial family-III cellulose-binding
  domain: a general mechanism for attachment to cellulose."
RL EMBO J. 15:5739-5751(1996).
RN [5]
RP X-RAY CRYSTALLOGRAPHY (1.7 ANGSTROMS) OF 1213-1361.
RX MEDLINE=98022914; PubMed=9402055;
RA Tavares G.A., Beguin P., Alzari P.M.;

```



```
FT DOMAIN 362 518 SUBSTRATE-BINDING (POTENTIAL).
FT DOMAIN 519 564 PRO/SER/THR-RICH (PT BOX).
FT DOMAIN 565 720 SUBSTRATE-BINDING (POTENTIAL).
FT DOMAIN 721 780 PRO/SER/THR-RICH (PT BOX).
FT DOMAIN 781 1331 CATALYTIC (ENDOGLUCANASE ACTIVITY).
FT ACT_SITE 162 162 PROTON DONOR (BY SIMILARITY).
FT ACT_SITE 257 257 NUCLEOPHILE (BY SIMILARITY).
FT CONFLICT 338 338 T -> P (IN REF. 2).
FT CONFLICT 340 346 TPTPTPT -> ROHOHQ (IN REF. 2).
SQ SEQUENCE 1331 AA: 146892 MW: 316892 DA: 146892
Query Match 42.5%; Score 352; DB 1; Length 1331;
Best Local Similarity 45.8%; Pred. No. 6.7e-25;
Matches 71; Conservative 30; Mismatches 50; Indels 4; Gaps 3;
OY 3 GGVKQVKNDSAPGDNQIKPGLQLVNTGSSVDLSLTVRYWFTRDGSSFLVYNCDA 62
DB 568 GQIKVLYANKETNSTNTIRPMLKVVNSGSSSIDLSRVTIRYWTVDGERAQSALS-DWA 626
OY 63 AMGCNIRASFSGVNPATPTADTYLQLSFTGCT--LAAGSGTGEIONRVKNSDWSNFDE 120
DB 627 QIGASNVTFKFKVLSSVSGADYLLIEIGFKSGAGQLQPKGTGEIQIREKNSDWSNYNOG 686
OY 121 NDYSY-GTNTAFQDMTKVTVYVNGRLVWGTEPSGT 154
DB 687 NDWSIQSMTSYGENEKEVTAYIDGVLVWGQEPSGT 721
RESULT 9
GUNA_CALSA STANDARD; PRT: 1742 AA.
ID GUNA_CALSA
AC P22534;
DT 01-AUG-1991 (Rel. 19, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE Endoglucanase A precursor (EC 3.2.1.4) (Endo-1,4-beta-glucanase A)
DE (Cellulase A).
GN CELA.
OS Caldocellum saccharolyticum (Caldicellulosiruptor saccharolyticus).
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Syntrophomonadaceae;
OC Caldicellulosiruptor.
OX NCBI_TaxID=44001;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=95336703; PubMed=7612247;
RA Te'o V.S., Saul D.J., Bergquist P.L.;
RT "celA, another gene coding for a multidomain cellulase from the
RT extreme thermophile Caldocellum saccharolyticum.";
RL Appl. Microbiol. Biotechnol. 43:291-296(1995).
RN [2]
RP SEQUENCE OF 1516-1742 FROM N.A.
RX MEDLINE=91247819; PubMed=2039230;
RA Leuthi E., Jasmat N.B., Grayling R.A., Love D.R., Bergquist P.L.;
RT "Cloning, sequence analysis, and expression in Escherichia coli of a
RT gene coding for a beta-mannanase from the extremely thermophilic
RT bacterium 'Caldocellum saccharolyticum'.";
RL Appl. Environ. Microbiol. 57:694-700(1991).
CC -1- FUNCTION: THE N-TERMINAL DOMAIN OF CELA ENCODES FOR AN
CC ENDOGLUCANASE ACTIVITY ON CARBOXYMETHYLCELLULOSE. THE C-TERMINAL
CC DOMAIN PROBABLY ACT SYNERGISTICALLY TO HYDROLYZE CRYSTALLINE
CC CELLULOSE.
CC -1- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic
CC linkages in cellulose.
CC -1- PTM: THE LINKER REGION (ALSO TERMED "HINGE") MAY BE A POTENTIAL
CC SITE FOR PROTEOLYSIS.
CC -1- SIMILARITY: IN THE N-TERMINAL SECTION; BELONGS TO CELLULASE FAMILY
CC E (FAMILY 9 OF GLYCOSYL HYDROLASES).
CC -1- SIMILARITY: IN THE C-TERMINAL SECTION; BELONGS TO CELLULASE FAMILY
CC L (FAMILY 48 OF GLYCOSYL HYDROLASES).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
```

```
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch)
CC -----
DR EMBL; L32742; AAA91086.1; -.
DR EMBL; M36063; AAA72860.1; -.
DR EMBL; L01257; -, NOT_ANNOTATED_CDS.
DR PIR; A43745; A43745.
DR HSSP; P26221; 1TF4.
DR InterPro; IPR001956; CBD_3.
DR InterPro; IPR001701; GH_9.
DR Pfam; IPR000556; Glyco_hydro_48.
DR Pfam; PF00759; Glyco_hydro_9; 1.
DR Pfam; PF00942; CBM_3; 3.
DR Pfam; PF02011; Glyco_hydro_48; 1.
DR PRINTS; PR00844; GLHYDRLASE48.
DR PRODOM; P001947; CBD_3; 2.
DR PRODOM; P001903; Glyco_hydro_48; 1.
DR PROSITE; PS00592; GLYCOSYL_HYDROL_F9_1; 1.
DR PROSITE; PS00598; GLYCOSYL_HYDROL_F9_2; 1.
KW Cellulose degradation; Hydrolase; Glycosidase; Signal; Repeat.
FT SIGNAL 1 23 POTENTIAL.
FT CHAIN 24 1742 ENDOGLUCANASE A.
FT DOMAIN 24 642 CATALYTIC 1.
FT DOMAIN 643 700 LINKER ("HINGE") (PRO-THR BOX).
FT DOMAIN 701 857 CELLULOSE-BINDING (BY SIMILARITY).
FT DOMAIN 858 903 LINKER ("HINGE") (PRO-THR BOX).
FT DOMAIN 904 1060 CELLULOSE-BINDING (BY SIMILARITY).
FT DOMAIN 1061 1112 LINKER ("HINGE") (PRO-THR BOX).
FT DOMAIN 1113 1742 CATALYTIC 2.
FT ACT_SITE 396 396 BY SIMILARITY.
FT ACT_SITE 434 434 BY SIMILARITY.
FT ACT_SITE 443 443 BY SIMILARITY.
FT CONFLICT 1545 1545 T -> A (IN REF. 2).
SQ SEQUENCE 1742 AA: 193696 MW: 316892 DA: 193696
Query Match 42.5%; Score 352; DB 1; Length 1742;
Best Local Similarity 45.5%; Pred. No. 9.2e-25;
Matches 71; Conservative 30; Mismatches 51; Indels 4; Gaps 3;
OY 2 SGVKVQYKNDSAPGDNQIKPGLQLVNTGSSVDLSLTVRYWFTRDGSSFLVYNCDA 61
DB 704 SQGIKVLVYANKETNSTNTIRPMLKVVNSGSSSIDLSRVTIRYWTVDGERAQSALS-DW 762
OY 62 AAMGCNIRASFSGVNPATPTADTYLQLSFTGCT--LAAGSGTGEIONRVKNSDWSNFDE 119
DB 763 AIGASNVTFKFKVLSSVSGADYLLIEIGFKSGAGQLQPKGTGEIQIREKNSDWSNYNO 822
OY 120 NDYSY-GTNTAFQDMTKVTVYVNGRLVWGTEPSGT 154
DB 823 GNDWSIQSMTSYGENEKEVTAYIDGVLVWGQEPSGT 858
RESULT 10
GUNA_CALSA STANDARD; PRT: 1039 AA.
ID GUNA_CALSA
AC P10474;
DT 01-JUL-1989 (Rel. 11, Created)
DT 01-JUL-1989 (Rel. 11, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Endoglucanase/exoglucanase B precursor [includes: Endoglucanase
DE (EC 3.2.1.4) (Endo-1,4-beta-glucanase) (Cellulase)
DE (Cellulohydrolase); Exoglucanase (EC 3.2.1.91) (Exocellulohydrolase)
DE (1,4-beta-cellobiohydrolase)].
GN CELB.
OS Caldocellum saccharolyticum (Caldicellulosiruptor saccharolyticus).
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Syntrophomonadaceae;
OC Caldicellulosiruptor.
OX NCBI_TaxID=44001;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=89098398; PubMed=2789517;
```

RA Saul D.J., Williams L.C., Love D.R., Chamley I.W., Bergquist P.I.;
 RT "Nucleotide sequence of a gene from *Caldocellum saccharolyticum*
 RL Nucleic Acids Res. 17:439-439(1989).
 CC -1- FUNCTION: THIS PROTEIN IS MADE UP OF TWO DOMAINS: THE N-TERMINAL
 CC DOMAIN HAS EXOGLUCANASE ACTIVITY WHILE THE C-TERMINAL DOMAIN IS
 CC AN ENDOGLUCANASE.
 CC -1- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic
 CC linkages in cellulose.
 CC -1- CATALYTIC ACTIVITY: Hydrolysis of 1,4-beta-D-glucosidic linkages
 CC in cellulose and cellobiose, releasing cellobiose from the non-
 CC reducing ends of the chains.
 CC -1- SIMILARITY: IN THE N-TERMINAL SECTION; BELONGS TO CELLULOSE FAMILY
 CC F (FAMILY 10 OF GLYCOSYL HYDROLASES).
 CC -1- SIMILARITY: IN THE C-TERMINAL SECTION; BELONGS TO CELLULOSE FAMILY
 CC A (FAMILY 5 OF GLYCOSYL HYDROLASES).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----

EMBL: X13602; CAA31936.1; .
 DR PIR: S02711; S02711.
 DR HSP: Q06851; INBC.
 DR InterPro: IPR001956; CBD_3.
 DR InterPro: IPR001547; GH_5.
 DR InterPro: IPR001000; Glyco_hydro_10.
 DR Pfam: PF00150; cellulase; 1.
 DR Pfam: PF00331; Glyco_hydro_10; 1.
 DR Pfam: PF00942; CBM_3; 1.
 DR PRINTS: PR00134; GLHYDRASE10.
 DR PRODOM: PD001947; CBD_3; 1.
 DR PROSITE: PS00591; GLYCOSYL_HYDROL_F10; 1.
 DR PROSITE: PS00659; GLYCOSYL_HYDROL_F5; 1.
 DR Cellulose degradation; Hydrolase; Glycosidase; Repeat;
 KW Multifunctional enzyme; Signal.
 FT SIGNAL 1 28
 FT CHAIN 29 1039 ENDOGLUCANASE/EXOGLUCANASE B.
 FT DOMAIN 376 416 THR/PRO-RICH, TANDM REPEATS OF T-P.
 FT DOMAIN 417 570 CELLULOSE-BINDING (BY SIMILARITY).
 FT DOMAIN 571 618 THR/PRO-RICH, TANDM REPEATS OF T-P.
 FT ACT_SITE 177 177 PROTON DONOR (POTENTIAL).
 FT ACT_SITE 285 285 NUCLEOPHILE (BY SIMILARITY).
 FT ACT_SITE 792 792 BY SIMILARITY.
 FT ACT_SITE 1039 AA; 117641 MW; OE0378171594DDAE CRC64;
 SQ SEQUENCE 1039 AA; 41.9%; Score 347; DB 1; Length 1039;

Query Match 41.9%; Score 347; DB 1; Length 1039;
 Best Local Similarity 45.5%; Pred. No. 1.5e-24;
 Matches 70; Conservative 30; Mismatches 50; Indels 4; Gaps 3;
 QY 3 CGVKVQYKNNDSAPCDNOIKPGLQLVNTGSSSVLDSTVTYVWFTRDGSSTLVYNCDA 62
 Db 420 GOIKVYANKENTSTNTIRPLWKLVNNGSSSIDLSRVTIRYVTVDDGSAQSAYS-DWA 478
 QY 63 AMGCNIRASFGSVNPATPTADTYLQLSFTGCT--LAAGSGTGEIQNRVKNKSDWSNFEET 120
 Db 479 QIGASNTVTFKVLSSSVSGADYLEICFKSGACGLQPKGTGEIQIRFNKSDWSNRYNQG 538
 QY 121 NDYSY-GTNTAFQDMTKVTVVYVNGRLVWGTEPSSG 153
 Db 539 NDWSLQMSYSGENKVTAVIDGVLVNGQBPSSG 572

RESULT 11

GUNW_ERWCA
 ID GUNW_ERWCA STANDARD; PRT; 504 AA.
 AC Q59395;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)

DT 01-NOV-1997 (Rel. 35, Last annotation update)
 DE Endoglucanase V1 precursor (EC 3.2.1.4) (Endo-1,4-beta-glucanase V1)
 DE (Cellulase V1).
 GN CELV1.
 OS Erwinia carotovora.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OC Pectobacterium.
 OX NCBI_TaxID=554;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=SCC3193;
 RX MEDLINE=95231512; PubMed=7715600;
 RT Mae A., Helkinhelmo R., Palva E.T.;
 RT "Structure and regulation of the Erwinia carotovora subspecies
 RT carotovora SCC3193 cellulase gene celv1 and the role of cellulase in
 RT phytopathogenicity";
 RL Mol. Gen. Genet. 247:17-26(1995).
 CC -1- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic
 CC linkages in cellulose.
 CC -1- SUBCELLULAR LOCATION: Secreted.
 CC -1- SIMILARITY: BELONGS TO CELLULOSE FAMILY A (FAMILY 5 OF GLYCOSYL
 CC HYDROLASES).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----

EMBL: X79241; CAA55823.1; .
 DR HSP: O85465; LA3H.
 DR InterPro: IPR001956; CBD_3.
 DR InterPro: IPR001547; GH_5.
 DR Pfam: PF00150; cellulase; 1.
 DR Pfam: PF00942; CBM_3; 1.
 DR PRODOM: PD001947; CBD_3; 1.
 DR PROSITE: PS00659; GLYCOSYL_HYDROL_F5; 1.
 DR Cellulose degradation; Hydrolase; Glycosidase; Signal.
 KW SIGNAL 1 31
 FT CHAIN 32 504 ENDOGLUCANASE V1.
 FT DOMAIN 32 334 CATALYTIC.
 FT DOMAIN 335 352 LINKER.
 FT DOMAIN 353 504 CELLULOSE-BINDING (BY SIMILARITY).
 FT ACT_SITE 168 168 PROTON DONOR (BY SIMILARITY).
 FT ACT_SITE 256 256 NUCLEOPHILE (BY SIMILARITY).
 SQ SEQUENCE 504 AA; 54963 MW; 0D7ECF74781565FA CRC64;

Query Match 41.4%; Score 343.5; DB 1; Length 504;
 Best Local Similarity 44.8%; Pred. No. 1.3e-24;
 Matches 69; Conservative 32; Mismatches 46; Indels 7; Gaps 4;

QY 2 SGVKVQYKNNDSAPCDNOIKPGLQLVNTGSSSVLDSTVTYVWFTRDGSSTLVYNC 59
 Db 354 TGDIVLQYRNVDNNDSDAIRMAFNKNTGSTPIKLSLDQVRYVPHDCKGPCANLFP--- 410
 QY 60 DWAAAGCCNIRASFGSVNPATPTADTYLQLSF-TGCTLAAGSGTGEIQNRVKNKSDWSNFD 118
 Db 411 DMWVGNPNIVTSTCTPAATDANKRNVLTFTFASGSLQPCAEVGTGEVQVRIHAGDSNVN 470
 QY 119 ETNDYSYCTN-TAFODTKVTVVYVNGRLVWGTEP 151
 Db 471 ETNDYSYGNITSYTNWDKITVHDKGTLVWGTEP 504

RESULT 12

GUNV_ERWCA
 ID GUNV_ERWCA STANDARD; PRT; 505 AA.
 AC Q47096;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 01-NOV-1997 (Rel. 35, Last annotation update)


```

ID AC GUNZ_CLOS STANDARD; PRT; 986 AA.
AC P23659;
DT 01-NOV-1991 (Rel. 20, Created)
DT 01-NOV-1991 (Rel. 20, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE Endoglucanase 2 precursor (EC 3.2.1.4) (Endo-1,4-beta-glucanase)
DE (Thermoactive cellulase) (Avicelase I).
GN CELZ.
OS Clostridium stercorarium.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1510;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 26-36 AND 475-486.
RC STRAIN=NCIB 11745;
RX MEDLINE=91066838; PubMed=2250652;
RA Jauris S., Ruecknagel K.P., Schwarz W.H., Kratzsch P.,
RA Bronnenmeier K., Staudenbauer W.L.;
RA "Sequence analysis of the Clostridium stercorarium celz gene encoding
a thermoactive cellulase (Avicelase I): Identification of catalytic
and cellulose-binding domains."
HL Mol. Gen. Genet. 223:258-267(1990).
CC -1- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic
linkages in cellulose.
CC -1- SIMILARITY: CONTAINS MULTIPLE CELLULOSE-BINDING SITES.
CC -1- SIMILARITY: BELONGS TO CELLULASE FAMILY E (FAMILY 9 OF GLYCOSYL
HYDROLASES).
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC
CC EMBL; X55299; CAA39010.1; ALT_SEQ.
DR PIR; S12021; S12021.
DR HSSP; P26221; 1TF4.
DR InterPro; IPR001956; CBD_3.
DR InterPro; IPR005102; DUF291.
DR InterPro; IPR001701; GH_9.
DR Pfam; PF00759; Glyco_hydro_9; 1.
DR Pfam; PF00942; CBM_3; 2.
DR Pfam; PF03442; DUF291; 2.
DR PRODOM; PD001947; CBD_3; 1.
DR PROSITE; PS00592; GLYCOSYL_HYDROL_F9_1; 1.
DR PROSITE; PS00698; GLYCOSYL_HYDROL_F9_2; 1.
KW Cellulose degradation; Hydrolase; Glycosidase; Repeat; Signal.
FT SIGNAL 1 25
FT CHAIN 26 986 ENDOGLUCANASE 2.
FT REPEAT 498 641 DOMAIN C'.
FT REPEAT 651 738 DOMAIN B'.
FT REPEAT 744 831 DOMAIN B'.
FT REPEAT 854 986 DOMAIN C.
FT DOMAIN 835 986 CELLULOSE-BINDING (BY SIMILARITY).
FT ACT_SITE 400 400 BY SIMILARITY.
FT ACT_SITE 438 438 BY SIMILARITY.
FT ACT_SITE 447 447 BY SIMILARITY.
SQ SEQUENCE 986 AA; 109512 MW; 1802E09B2292323690 CRC64;

Query Match 34.4%; Score 285; DB 1; Length 986;
Best Local Similarity 37.7%; Pred. No. 8,le-19;
Matches 58; Conservative 36; Mismatches 52; Indels 8; Gaps 3;

Qy 2 SGGVKVQYKKNDSAPGDNQIKPGLQLVNTGSSVDLSLTVTVRYWFTRDGGSTLVYNCW 61
Db 837 TGVLOIQMFNTRSDKTINGIPRYRLTNGTTPRLSDVKIRYYITDGEKDNQFW-CDW 895

Qy 62 AAMCGNTRATSGVSNPATPTADTYLQLSFTGCGTLAAG----GSGTQNRVKNKSDWSNF 117
Db 896 SSVGSNNTITGFVKMAEPKEGADYYLETFGTGTD----AGYLPQNSIEVQNRFSKADWTDY 952

```

```

Qy 118 DETNDYSYGTNTAFQDWTKTVTYVVGKGLVMGTPEP 151
Db 953 IOTNDYSFSTNTSYGSNDRIITYISGVLSGIEP 986

RESULT 15
GUNI_CLOTM
ID GUNI_CLOTM STANDARD; PRT; 879 AA.
AC Q02934;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Endoglucanase I precursor (EC 3.2.1.4) (EG1) (Endo-1,4-beta-glucanase)
DE (Cellulase I).
GN CELI.
OS Clostridium thermocellum.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1515;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 56-69.
RC STRAIN=NCIB 10682;
RX MEDLINE=93171873; PubMed=8436949;
RA Hazlewood G.P., Davidson K., Laurie J.I., Huskisson N.S.,
RA Gilbert H.J.;
RT "Gene sequence and properties of Celli, a family E endoglucanase from
Clostridium thermocellum."
RL J. Gen. Microbiol. 139:307-316(1993).
CC -1- FUNCTION: THIS ENZYME CATALYZES THE ENDOHYDROLYSIS OF 1,4-HEXA-
GLUCOSIDIC LINKAGES IN CELLULOSE, LICHENIN AND CEREAL BETA-D-
GLUCANS. PRINCIPALLY ACTIVE AGAINST BARLEY BETA-GLUCAN.
CC -1- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic
linkages in cellulose.
CC -1- PATHWAY: Cellulose degradation.
CC -1- SIMILARITY: BELONGS TO CELLULASE FAMILY E (FAMILY 9 OF GLYCOSYL
HYDROLASES).
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC
CC EMBL; L04735; AAA20892.1;
DR PIR; A47704; A47704.
DR HSSP; P26221; 1TF4.
DR InterPro; IPR001956; CBD_3.
DR InterPro; IPR001701; GH_9.
DR Pfam; PF00759; Glyco_hydro_9; 1.
DR Pfam; PF00942; CBM_3; 2.
DR PRODOM; PD001947; CBD_3; 1.
DR PROSITE; PS00592; GLYCOSYL_HYDROL_F9_1; 1.
DR PROSITE; PS00698; GLYCOSYL_HYDROL_F9_2; 1.
KW Cellulose degradation; Hydrolase; Glycosidase; Signal.
FT SIGNAL 1 55
FT CHAIN 56 879 ENDOGLUCANASE 1.
FT DOMAIN 56 518 CATALYTIC.
FT DOMAIN 729 879 CELLULOSE-BINDING (BY SIMILARITY).
FT ACT_SITE 448 448 BY SIMILARITY.
FT ACT_SITE 486 486 BY SIMILARITY.
FT ACT_SITE 495 495 BY SIMILARITY.
SQ SEQUENCE 879 AA; 97796 MW; 35A60069A514A927 CRC64;

Query Match 31.0%; Score 257; DB 1; Length 879;
Best Local Similarity 37.8%; Pred. No. 2,8e-16;
Matches 48; Conservative 34; Mismatches 41; Indels 4; Gaps 3;

```

```

Qy 1 VSGGVKVOYKKNDSAPGDNQIKPGLQLVNTGSSVDLSLTVTVRYWFTRDGGSTLVYNCW 60
Db 736 IKGEVLYQVANGNAGATSNINPKIINNGTKAINLSVDKIRYYTYTKEGGASQNF-CD 794

```

Search completed: November 13, 2002, 11:55:04
Job time : 4.54023 secs


```
Db 79 QVGCNVRGKSFVYKLTGRTGADYYIEITFTSGAGSLAAGSSGDIQVRIKNDWNTYNEA 138
QY 121 NDYSYG-TNTAFQDMTKVTYVYNGRLVWCTEP 151
DB 139 NDYSYDPTKTSFADWNRVTLRNGQLVWVEP 170

RESULT 2
Q9L8L8
ID Q9L8L8 PRELIMINARY; PRT; 921 AA.
AC Q9L8L8;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Beta-1,4-xylanase Xyna precursor.
GN Xyna.
OS Caldicellulosus cellulovorans.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;
OC Allicyclobacillaceae; Caldicellulosus.
OX NCBI_TaxID=74586;
KW SIGNAL.
FT CHAIN 1 33 POTENTIAL.
SQ SEQUENCE 921 AA: 102380 MW: C5DDDIAT7F567413 CRC64;

Query Match 51.7%; Score 429; DB 2; Length 921;
Best Local Similarity 51.6%; Pred. No. 6.7e-27;
Matches 79; Conservative 35; Mismatches 37; Indels 2; Gaps 2;

QY 2 SGCVKQVYKNDSDAPGDNQIKPGLQVLVNTGSSVDLSTVTYVYFTRDGGSSTLVYNCDA 61
DB 563 SCTLKVEYRVGDSSTDNQMKPOLRIVNTGSOAVPLTELKVRWYTKN-STQAEQYFCDW 621

QY 62 AAMGCGNIRASFGSNVNPATPTADTYLQLSFTGTLAAGSGTGEIONRVNKSDFDET 121
DB 622 AQICGSNIRAFVSLAQPVSGADSVIELSFTGSGVPAGNTEIONRIHFTNMMNINETD 681

QY 122 DYSY-GTNTAFQDMTKVTYVYNGRLVWCTEP 153
DB 682 DWSYNGTOTTWGPSTRITLYRNGVLVWCTEP 714

RESULT 3
Q9RFX5
ID Q9RFX5 PRELIMINARY; PRT; 930 AA.
AC Q9RFX5;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Multidomain beta-1,4-mannanase precursor.
GN MANA.
OS Caldicellulosus cellulovorans.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;
```

```
OC Allicyclobacillaceae; Caldicellulosus.
OX NCBI_TaxID=74586;
RN 11;
RP SEQUENCE FROM N.A.
RX MEDLINE=20120520; PubMed=10653733;
RA Sunna A., Gibbs M.D., Chin C.W.J., Nelson P.J., Bergquist P.L.;
RT "A gene encoding a novel multidomain beta-1,4-mannanase from
RT Caldicellulosus cellulovorans and action of the recombinant enzyme on
RT kraft pulp.";
RL Appl. Environ. Microbiol. 66:664-670(2000).
DR EMBL; AF163837; AAF22274.1; -.
DR HSSP; Q06851; INBC.
DR InterPro; IPR001956; CBD_3.
DR InterPro; IPR004302; Chitin_binding_3.
DR InterPro; IPR001547; GH_5.
DR InterPro; IPR002965; P_Rich_extensn.
DR Pfam; PF00942; CBM_3; 2.
DR Pfam; PF00150; cellulase; 1.
DR Pfam; PF03067; Chitin_bind_3; 1.
DR PRINTS; PR01217; PRICHEXTENS.
DR ProDom; PD001947; CHD_3; 2.
DR ProSITE; PS00659; GLYCOSYL_HYDROL_F5; UNKNOWN_1.
KW SIGNAL.
FT CHAIN 1 33 POTENTIAL.
SQ SEQUENCE 930 AA: 101576 MW: 0086638D54DIA2CC CRC64;

Query Match 51.7%; Score 429; DB 2; Length 930;
Best Local Similarity 51.3%; Pred. No. 6.7e-27;
Matches 78; Conservative 31; Mismatches 39; Indels 4; Gaps 3;

QY 3 GGVKQVYKNDSDAPGDNQIKPGLQVLVNTGSSVDLSTVTYVYFTRDGGSSTLVYNCDA 62
DB 780 GNLVQVYRAADTNATDNQKLPFRIVNCRCTSSVPLSELTYWYTVYD-GDKQVFNCDWA 838

QY 63 AAMGCGNIRASFGSNVNPATPTADTYLQLSFT--GGTLAAGSGTGEIONRVNKSDFDET 120
DB 839 QVGCNLRGKSFVYKLTGRTGADYYIEITFTSGAGSLAAGSSGDIQVRIKNDWNTYNEA 898

QY 121 NDYSYG-TNTAFQDMTKVTYVYNGRLVWCTEP 151
DB 899 NDYSYDPTKTSFADWNRVTLRNGQLVWVEP 930

RESULT 4
Q9Z4I1
ID Q9Z4I1 PRELIMINARY; PRT; 997 AA.
AC Q9Z4I1;
DT 01-MAY-1999 (TrEMBLrel. 10, Created)
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Cellulase precursor (EC 3.2.1.4).
GN CELB.
OS Bacillus sp. BP-23.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;
OC Bacillaceae; Bacillus.
OX NCBI_TaxID=89769;
RN 11;
RP SEQUENCE FROM N.A.
RX STRAIN=BP-23;
RX MEDLINE=21129642; PubMed=11234960;
RA Pastor F.I.J., Pujol X., Blanco A., Vidal T., Torres A.L., Diaz P.;
RT "Molecular cloning and characterization of a multidomain endoglucanase
RT from Paenibacillus sp BP-23: evaluation of its performance in pulp
RT refining.";
RL Appl. Microbiol. Biotechnol. 55:61-68(2001).
CC -1- SIMILARITY: CONTAINS 2 FIBRONECTIN TYPE III-LIKE DOMAINS.
DR EMBL; AJ133614; CAB38941.1; -.
DR HSSP; P26221; 1TF4.
DR InterPro; IPR001956; CBD_3.
DR InterPro; IPR003961; FN_III.
DR InterPro; IPR001701; GH_9.
DR InterPro; IPR000566; Lipoclin_cytFABP.
```


RESULT 6		PRELIMINARY:		PRT:		508 AA.	
Q93LDO	Q93LDO						
AC	Q93LDO						
DT	01-DEC-2001 (TrEMBLrel. 19, Created)						
DT	01-DEC-2001 (TrEMBLrel. 19, Last sequence update)						
DT	01-JUN-2002 (TrEMBLrel. 21, Last annotation update)						
DE	Endo-1,4-beta-glucanase (EC 3.2.1.4).						
OS	Bacillus subtilis.						
OC	Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;						
OC	Bacillaceae; Bacillus.						
OX	NCBI_TaxID=1423;						
RN	[1]						
RP	SEQUENCE FROM N.A.						
RC	STRAIN=CH21;						
RT	Zvidzai C.J., Delgado O.D., Zvaunya R., Mattiasson B.;						
RT	"Bacillus subtilis CH21 endo-b-1,4-glucanase gene.";						
RL	Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.						
DR	EMBL: AY044252; AAK94871.1;						
DR	InterPro: IPR001956; CBD_3.						
DR	InterPro: IPR001547; GH_5.						
DR	Pfam: PF00942; CBM_3; 1.						
DR	Pfam: PF00150; cellulase; 1.						
DR	ProDom: PD001947; CBD_3; 1.						
DR	PROSITE: PS00659; GLYCOSYL_HYDROL_F5; UNKNOWN_1.						
KW	Glycosidase; Hydrolase.						
SQ	SEQUENCE 508 AA; 56553 MW; A395EEIC6F3340F2 CRG64;						
<p>Query Match 44.6%; Score 370; DB 2; Length 508; Best Local Similarity 44.0%; Pred. No. 2.2e-22; Matches 66; Conservative 34; Mismatches 46; Indels 4; Gaps</p>							
Qy	4 GVKVQYKNNDSAPGDNOIKPGLQLVNTGSSVDLSTVTVRYWF-TRDGGSTLVYNCDDWA 62						
Db	362 GISVQYRAGDGSNNSNOIRPOLQIKNGNTVLDKDYARYKAKNGQON---FDCDYA 418						
Qy	63 AMCGCNTRASFGSPATPADTYLQLSFTGGTAAAGSGTGETQNRVNKSDMSNFQETND 122						
Db	419 QICGCVNTHKEVTLHKPKQADTYLELGFKNGLAPCASTGNLQLRLHNDHRSNYAOSGD 478						
Qy	123 YSGTNTAFODWTKVTVYVNGRLVWGTEPS 152						
Db	479 YSPFKSNTEFTTKKITLYDQGLIWGTEPN 508						
<p>RESULT 7</p>							
Q52731	Q52731						
ID	Q52731						
AC	Q52731						
DT	01-JUN-1998 (TrEMBLrel. 06, Created)						
DT	01-JUN-1998 (TrEMBLrel. 06, Last sequence update)						
DT	01-JUN-2002 (TrEMBLrel. 21, Last annotation update)						
DE	Endo-b-1,4-glucanase (EC 3.2.1.4).						
GN	CELS.						
OS	Bacillus sp. 79-23.						
OC	Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;						
OC	Bacillaceae; Bacillus.						
OX	NCBI_TaxID=72363;						
RN	[1]						
RP	SEQUENCE FROM N.A.						
RC	STRAIN=79-23;						
RC	Jung K.-H., Chun Y.C., Lee J.-C., Kim J.H., Yoon K.-H.;						
RL	Submitted (JAN-1998) to the EMBL/GenBank/DBJ databases.						
RL	EMBL: AF045482; AAC02536.1;						
DR	HSSP: 085465; 1A3H.						

```
DR InterPro: IPR001956; CBD_3.
DR InterPro: IPR001547; GH_5.
DR Pfam: PF00942; CBM_3; 1.
DR Pfam: PF00150; cellulase; 1.
DR ProDom: PD001947; CBD_3; 1.
DR PROSITE: PS00659; GLYCOSYL_HYDROL_F5; UNKNOWN_1.
KW Glycosidase; Hydrolase.
SQ SEQUENCE 499 AA; 55193 MW; 990A1878788CF738 CRC64;

Query Match 43.8%; Score 363; DB 2; Length 499;
Best Local Similarity 43.38; Pred. No. 8e-22;
Matches 65; Conservative 33; Mismatches 48; Indels 4; Gaps 2;

Qy 4 GVKVOYKNDSAPGDNQIKPGLQLVNTGSSVDLSTVTYRYWF-TRDGGSTLVYNCOWA 62
Db 353 GTSVQYRAGDCGSMNSQIRPOLKIKNNGNTVLDKDVTRYWYKAKNGQN---DCDYA 409

Qy 63 AMGCNIRASFGSVNTPATPTADTYLQLSFTGCTLAAGSTGEIQNRVKNKSDNSNFDEND 122
Dy 410 QIGCGNLTHKEVTLHKPKQGADTYLELFGKNTLPGPGASTGNIQLRLHNDWNSYAOQSGD 469

Qy 123 YSYGTNTAFQDWTKVTVVYVNGRLVWGTEPS 152
Dy 470 YSFFQSNTEFTTKKITLYDQCKLINGTEPN 499

RESULT 8
Q45532 PRELIMINARY; PRT; 499 AA.
AC Q45532;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Cellulase.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;
OC Bacillaceae; Bacillus.
OX NCBI_TaxID=1423;

RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE=87190397; PubMed=3106035;
RX Nakamura A., Uozumi T., Beppu T.;
RT "Nucleotide sequence of a cellulase gene of Bacillus subtilis.";
RL Eur. J. Biochem. 164:317-320(1987).
DR EMBL; M28332; AAA22307.1; -.
DR HSSP; O85465; 1A3H.
DR InterPro: IPR001956; CBD_3.
DR InterPro: IPR001547; GH_5.
DR Pfam: PF00942; CBM_3; 1.
DR Pfam: PF00150; cellulase; 1.
DR ProDom: PD001947; CBD_3; 1.
DR PROSITE; PS00659; GLYCOSYL_HYDROL_F5; UNKNOWN_1.
SQ SEQUENCE 499 AA; 55075 MW; D424AB9E53B94D23 CRC64;

Query Match 43.4%; Score 360; DB 2; Length 499;
Best Local Similarity 43.68; Pred. No. 1.4e-21;
Matches 65; Conservative 32; Mismatches 48; Indels 4; Gaps 2;

Qy 4 GVKVOYKNDSAPGDNQIKPGLQLVNTGSSVDLSTVTYRYWF-TRDGGSTLVYNCOWA 62
Dy 353 GTSVQYRAGDCGSMNSQIRPOLKIKNNGNTVLDKDVTRYWYKAKNGQN---PCDYA 409

Qy 63 AMGCNIRASFGSVNTPATPTADTYLQLSFTGCTLAAGSTGEIQNRVKNKSDNSNFDEND 122
Dy 410 QIGCGNLTHKEVTLHKPKQGADTYLELFGKNTLPGPGASTGNIQLRLHNDWNSYAOQSGD 469

Qy 123 YSYGTNTAFQDWTKVTVVYVNGRLVWGTEP 151
Dy 470 YSFFQSNTEFTTKKITLYHQCKLINGTEP 498

RESULT 9
P96311 PRELIMINARY; PRT; 1711 AA.
AC P96311;
DT 01-MAY-1997 (TrEMBLrel. 03, Created)
DT 01-MAY-1997 (TrEMBLrel. 03, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Endoglucanase A (EC 3.2.1.4) (Endo-1,4-beta-glucanase A) (Cellulase A) (Fragment).
GN CELA.
OS Anaerocellum thermophilum.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridia;
OC Clostridiales; Anaerocellum group; Anaerocellum.
OX NCBI_TaxID=31899;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN=2-1320;
RA Zverlov V.;
RL Submitted (MAR-1997) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: THE N-TERMINAL DOMAIN OF CELA ENCODES FOR AN
CC ENDOLUCANASE ACTIVITY ON CARBOXYMETHYLCELLULOSE. THE C-TERMINAL
CC DOMAIN PROBABLY ACT SYNERGISTICALLY TO HYDROLYZE CRYSTALLINE
CC CELLULOSE (BY SIMILARITY).
CC -!- CATALYTIC ACTIVITY: ENDOHYDROLYSIS OF 1,4-BETA-D-GLUCOSIDIC
CC LINKAGES IN CELLULOSE.
CC -!- SIMILARITY: THE N-TERMINAL PART BELONGS TO CELLULOSE FAMILY E
CC (FAMILY 9 OF GLYCOSYL HYDROLASES).
CC -!- SIMILARITY: THE C-TERMINAL PART BELONGS TO CELLULOSE FAMILY L
CC (FAMILY 48 OF GLYCOSYL HYDROLASES).
DR EMBL; Z86105; CAB06786.1; -.
DR HSSP; Z62221; 1TF4.
DR InterPro: IPR001956; CBD_3.
DR InterPro: IPR001701; GH_9.
DR InterPro: IPR000556; Glyco_hydro_48.
DR Pfam: PF00942; CBM_3; 3.
DR Pfam: PF02011; Glyco_hydro_48; 1.
DR Pfam: PF00759; Glyco_hydro_9; 1.
DR PRINTS; PR00844; GLHYDRLASE48.
DR ProDom: PD001947; CBD_3; 2.
DR ProDom: PD011903; Glyco_hydro_48; 1.
DR PROSITE; PS00592; GLYCOSYL_HYDROL_F9_1; 1.
DR PROSITE; PS00698; GLYCOSYL_HYDROL_F9_2; 1.
KW Cellulose degradation; Hydrolase; Glycosidase; Signal; Repeat.
FT NON_TER 1
SQ SEQUENCE 1711 AA; 189979 MW; E3E987CEB9CD0C21 CRC64;

Query Match 43.4%; Score 359.5; DB 2; Length 1711;
Best Local Similarity 46.68; Pred. No. 6.6e-21;
Matches 76; Conservative 25; Mismatches 45; Indels 17; Gaps 5;

Qy 1 VSGG-VKVOYKNDSAPGDNQIKPGLQIVNTGSSVDLSTVTYRYWFTHDGSSSTLVYNC 59
Dy 882 VAGQIKVLYANKETNSTTIRPWLKVVNTGSSSIDLSRVTIRYWTVDGKAQSAIS- 940

Qy 60 DWAAMGCCNIRASFGSVNTPATPTADTYLQLSFTGCT--LAAGSTGEIQNRVKNKSDNSNF 117
Dy 941 DWAQIGASNVTFKPVKLVSSVSGADYILEIGFKSAGAGLQAGKDTGQTQIRPNKSDNSNY 1000

Qy 118 DETNDYS-----YGTNTAFQDWTKVTVVYVNGRLVWGTEPSG 153
Dy 1001 NQGDWNSHWQSNWTNYGEN-----VKVTAYIDGVLVWGTEPSG 1037

RESULT 10
O83012 PRELIMINARY; PRT; 501 AA.
AC O83012;
DT 01-NOV-1998 (TrEMBLrel. 08, Created)
DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Cellulase.
OS Bacillus sp.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;
OC Bacillaceae; Bacillus.
OX NCBI_TaxID=1409;
```



```

DE XyNA.
GN XyNA.
OS Caldicellulosiraptor sp. Tok7B.1.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridia;
OC Clostridiales; Syntrophomonadaceae; Caldicellulosiraptor.
OX NCBI_TaxID=80339;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Tok7B.1.
RX MEDLINE=20171169; PubMed=10706665;
RA Gibbs M.D., Reeves R.A., Farrington G.K., Anderson P., Williams D.P.,
RA Bergquist P.L.;
RT "Multidomain and multifunctional glycosyl hydrolases from the extreme
RT thermophile Caldicellulosiraptor isolate Tok7B.1.";
RL Curr. Microbiol. 40:333-340(2000).
DR EMBL: AF078737; AD30363.1; -
DR HSSP: Q06851; INBC.
DR InterPro: IPR001956; CBD_3.
DR InterPro: IPR005084; CBM_6.
DR InterPro: IPR003305; CBM_CenC.
DR InterPro: IPR001000; Glyco_hydro_10.
DR InterPro: IPR002965; P_rich_extensn.
DR Pfam: PF00942; CBM_3; 3.
DR Pfam: PF02018; CBM_4_9; 2.
DR Pfam: PF03422; CBM_6; 1.
DR Pfam: PF00331; Glyco_hydro_10; 1.
DR PRINTS: PR00134; GLHYDRLASE10.
DR PRINTS: PR01217; PRICHEXTENS.
DR ProDom: PD001947; CBD_3; 3.
DR PROSITE: PS00591; GLYCOSYL_HYDROL_F10; 1.
SQ SEQUENCE 1770 AA; 193641 MW; 8BAF1937D4926C92 CRC64;

Query Match 41.7%; Score 346; DB 2; Length 1770;
Best Local Similarity 44.0%; Pred. No. 8.7e-20;
Matches 70; Conservative 27; Mismatches 46; Indels 16; Gaps 4;

QY 4 GVKVQYKNNDSAPGDNQIKPGLQLVNTGSSVSLSTVTYRYWFTRDGSSSTLVYNCDA 63
   I:II IIIII::: : I:II IIIIIII I:IIII I I IIII
Db 1104 GLKVLKNNETSASTSIRPWFKIVNGSSVSLSRVKIRIYWTVDGKPOSAY-CDWAQ 1162

QY 64 MCGCNIRASFGSVNPATPTADTYLQLSFTGCT--LAAGSGTGEIQNRVKNKSDWSNFD 121
   I:II IIIII::: : I:II IIIIIII I:IIII I I IIII
Db 1163 IGASNVTFNFVKLSGSGVADYILEVGFSSGAGLOPKDGTGDIQVRFNKNDWSNYNQAD 1222

QY 122 DYS-----YGTNTAFQDWTKVTVYVNGRLVWGTEPSG 153
   I:II IIIII::: : I:II IIIIIII I:IIII I I IIII
Db 1223 DWSLQSMTDYGEN-----AKVTLYVDGVLVWGQEPGG 1255

RESULT 14
Q9AQH0 PRELIMINARY; PRT; 996 AA.
AC Q9AQH0
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 21, Last annotation update)
DE Glycosyl hydrolase 5 (Fragment).
OS Caldicellulosiraptor sp. Tok7B.1.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridia;
OC Clostridiales; Syntrophomonadaceae; Caldicellulosiraptor.
OX NCBI_TaxID=80339;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Tok7B.1.
RX MEDLINE=20171169; PubMed=10706665;
RA Gibbs M.D., Reeves R.A., Farrington G.K., Anderson P., Williams D.P.,
RA Bergquist P.L.;
RT "Multidomain and multifunctional glycosyl hydrolases from the extreme
RT thermophile caldicellulosiraptor isolate Tok7B.1.";
RL Curr. Microbiol. 40:333-340(2000).
DR EMBL: AF078038; AAK06388.1; -
DR HSSP: Q06851; INBC.
DR InterPro: IPR001956; CBD_3.

```

```

DR InterPro: IPR002860; GH_BNR.
DR InterPro: IPR002965; P_rich_extensn.
DR Pfam: PF02012; BNR; 9.
DR Pfam: PF00942; CBM_3; 1.
DR PRINTS: PR01217; PRICHEXTENS.
DR ProDom: PD001947; CBD_3; 1.
KW Hydrolase.
KW NON_TER.
FT NON_TER.
SQ SEQUENCE 996 AA; 108275 MW; 3C72B6ED22F3C614 CRC64;

Query Match 41.5%; Score 344; DB 2; Length 996;
Best Local Similarity 44.0%; Pred. No. 6.4e-20;
Matches 70; Conservative 26; Mismatches 47; Indels 16; Gaps 4;

QY 4 GVKVQYKNNDSAPGDNQIKPGLQLVNTGSSVSLSTVTYRYWFTRDGSSSTLVYNCDA 63
   I:II IIIII::: : I:II IIIIIII I:IIII I I IIII
Db 844 GLKVLKNNETSASAGSIRPWFKIVNGSSVSLSRVKIRIYWTVDGKPOSAY-CDWAQ 902

QY 64 MCGCNIRASFGSVNPATPTADTYLQLSFTGCT--LAAGSGTGEIQNRVKNKSDWSNFD 121
   I:II IIIII::: : I:II IIIIIII I:IIII I I IIII
Db 903 IGASNVTFNFVKLSGSGVADYILEVGFSSGAGLOPKDGTGDIQVRFNKNDWSNYNQAD 962

QY 122 DYS-----YGTNTAFQDWTKVTVYVNGRLVWGTEPSG 153
   I:II IIIII::: : I:II IIIIIII I:IIII I I IIII
Db 963 DWSLQSMTDYGEN-----AKVTLYVDGVLVWGQEPGG 995

RESULT 15
O52374 PRELIMINARY; PRT; 1779 AA.
AC O52374
DT 01-JUN-1998 (TrEMBLrel. 06, Created)
DT 01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Family 10 xylanase (EC 3.2.1.8).
GN XYN.
OS Caldicellulosiraptor sp. Rt69B.1.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridia;
OC Clostridiales; Syntrophomonadaceae; Caldicellulosiraptor.
OX NCBI_TaxID=70295;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Rt69B.1.
RA Morris D.D., Gibbs M.D., Ford M., Thomas J., Bergquist P.L.;
RA Family 10 and 11 xylanase genes from Caldicellulosiraptor sp.
RT Rt69B.1.;
RL Submitted (DEC-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF036924; AAB95326.1; -
DR HSSP: Q06851; INBC.
DR InterPro: IPR001956; CBD_3.
DR InterPro: IPR005084; CBM_6.
DR InterPro: IPR003305; CBM_CenC.
DR InterPro: IPR001000; Glyco_hydro_10.
DR Pfam: PF00942; CBM_3; 3.
DR Pfam: PF02018; CBM_4_9; 2.
DR Pfam: PF03422; CBM_6; 1.
DR Pfam: PF00331; Glyco_hydro_10; 1.
DR PRINTS: PR00134; GLHYDRLASE10.
DR ProDom: PD001947; CBD_3; 3.
DR PROSITE: PS00591; GLYCOSYL_HYDROL_F10; 1.
DR Glycosidase; Hydrolase; Xylan degradation.
KW Glycosidase.
SQ SEQUENCE 1779 AA; 194304 MW; CE5269B8606B5CED CRC64;

Query Match 41.5%; Score 344; DB 2; Length 1779;
Best Local Similarity 44.0%; Pred. No. 1.3e-19;
Matches 70; Conservative 25; Mismatches 48; Indels 16; Gaps 4;

QY 4 GVKVQYKNNDSAPGDNQIKPGLQLVNTGSSVSLSTVTYRYWFTRDGSSSTLVYNCDA 63
   I:II IIIII::: : I:II IIIIIII I:IIII I I IIII
Db 1113 GLKVLKNNETSASTSIRPWFKIVNGSSVSLSRVKIRIYWTVDGKPOSAY-CDWAQ 1171

QY 64 MCGCNIRASFGSVNPATPTADTYLQLSFTGCT--LAAGSGTGEIQNRVKNKSDWSNFD 121
   I:II IIIII::: : I:II IIIIIII I:IIII I I IIII

```

Db 1172 ICASNVT'NFVKLTSGVSGADYYLEVGFSSGACQLQPGKDTGDIQVR'FNKNDWSNYNQAD 1231

Qy 122 DYS-----YGTNTAFQDWTKVTVVYVNGRLVWGTEPSG 153

Ub 1232 DWSWMQSMNTNYGEN-----AKVTLYVDGLVWVGQEPGG 1264

Search completed: November 13, 2002, 11:54:11
Job time : 16.2716 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run On: November 13, 2002, 11:49:46 ; Search time 4.75718 Seconds
(without alignments)
952.482 Million cell updates/sec

Title: US-09-917-378-4

Perfect score: 829

Sequence: 1 VSGGVKVOYKNNNSAPGDNQ.....TKVTYYVNGRLVWGTEPSGT 154

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA.*

1: /cgn2_6/ptodata/1/1aa/5A_COMB.pep.*
2: /cgn2_6/ptodata/1/1aa/5B_COMB.pep.*
3: /cgn2_6/ptodata/1/1aa/6A_COMB.pep.*
4: /cgn2_6/ptodata/1/1aa/6B_COMB.pep.*
5: /cgn2_6/ptodata/1/1aa/PCTUS_COMB.pep.*
6: /cgn2_6/ptodata/1/1aa/backfiles1.pep.*

Prod. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	374	45.1	551	2	US-09-033-537A-1
2	373.5	45.1	700	2	US-07-862-588B-2
3	357.5	43.1	493	4	US-09-198-956-10
4	357.5	43.1	493	4	US-09-198-955A-12
5	357.5	43.1	493	4	US-09-694-531-12
6	357.5	43.1	493	4	US-09-670-141-10
7	356.5	43.0	167	5	PCT-US95-13813-9
8	349	42.1	616	4	US-09-136-574A-47
9	348	42.0	1751	4	US-09-136-574A-44
10	340	41.0	1426	4	US-09-136-574A-43
11	226	27.3	162	1	US-08-048-164A-2
12	226	27.3	162	1	US-08-460-462-2
13	226	27.3	162	1	US-08-460-457-2
14	226	27.3	162	1	US-08-460-458-2
15	226	27.3	162	2	US-08-460-455-2
16	226	27.3	162	2	US-08-330-394A-2
17	226	27.3	163	3	US-09-006-636-7
18	226	27.3	163	4	US-09-006-632-7
19	226	27.3	163	4	US-09-325-274-7
20	225	27.1	382	4	US-09-277-716-22
21	225	27.1	382	4	US-09-609-161B-22
22	224	27.0	156	2	US-08-330-394A-22
23	218	26.3	154	2	US-08-330-394A-29
24	127	15.3	531	2	US-07-862-588B-7
25	85	10.3	428	3	US-09-118-319-5
26	85	10.3	2123	4	US-08-968-685A-10
27	84	10.1	434	2	US-09-008-962-1

Query Match 45.1% ; Score 374 ; DB 2 ; Length 551 ;

28 84 10.1 434 2 US-08-675-507-1 Sequence 1, Appli
29 84 10.1 434 3 US-09-213-205-1 Sequence 1, Appli
30 84 10.1 1833 4 US-08-621-944A-4 Sequence 4, Appli
31 84 10.1 1833 4 US-08-945-567D-4 Sequence 4, Appli
32 84 10.1 1833 4 US-08-621-944A-3 Sequence 3, Appli
33 84 10.1 1992 4 US-08-945-567D-3 Sequence 3, Appli
34 82 9.9 2048 4 US-09-268-347-48 Sequence 48, Appli
35 80 9.7 326 4 US-09-286-691-23 Sequence 23, Appli
36 80 9.7 326 4 US-09-687-147-23 Sequence 23, Appli
37 80 9.7 1338 2 US-08-728-470-9 Sequence 9, Appli
38 80 9.7 1338 4 US-08-719-641-9 Sequence 9, Appli
39 80 9.7 1529 2 US-08-728-470-10 Sequence 10, Appli
40 80 9.7 1529 4 US-08-719-641-10 Sequence 10, Appli
41 80 9.7 1599 2 US-08-617-697-9 Sequence 9, Appli
42 80 9.7 1600 2 US-08-617-697-10 Sequence 10, Appli
43 79.5 9.6 333 3 US-08-988-111-3 Sequence 3, Appli
44 79.5 9.6 333 4 US-09-387-922-3 Sequence 3, Appli
45 79.5 9.6 892 4 US-09-336-447A-5 Sequence 5, Appli

ALIGNMENTS

RESULT 1
US-09-033-537A-1
: Sequence 1, Application US/09033537A
: Patent No. 5958083
: GENERAL INFORMATION:
: APPLICANT: Onishi, Masahiro
: APPLICANT: Fich, Merete
: APPLICANT: Toft, Annette Hanne
: APPLICANT: Shlein, Martin
: TITLE OF INVENTION: Prevention Of Back-Staining
: TITLE OF INVENTION: In Stone Washing
: NUMBER OF SEQUENCES: 1
: CORRESPONDENCE ADDRESS:
: ADDRESSER: No. 5958083o No. 5958083disk of No. 5958083th America, Inc.
: STREET: 405 Lexington Avenue
: CITY: New York
: STATE: NY
: COUNTRY: U.S.A.
: ZIP: 10174
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Diskette
: COMPUTER: IBM Compatible
: OPERATING SYSTEM: DOS
: SOFTWARE: Fastseq for Windows Version 2.0
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/09/033.537A
: FILING DATE: 02-MAR-1998
: CLASSIFICATION: 008
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: 0993/95
: FILING DATE: 08-SEP-1995
: APPLICATION NUMBER: PCT/DK96/00364
: FILING DATE: 03-SEP-1996
: ATTORNEY/AGENT INFORMATION:
: NAME: Green, Reza
: REGISTRATION NUMBER: 38.475
: REFERENCE/DOCKET NUMBER: 4492.204-US
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 212-867-0123
: TELEFAX: 212-878-9655
: TELEX:
: INFORMATION FOR SEQ ID NO: 1:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 551 amino acids
: TYPE: amino acid
: STRANDEDNESS: single
: TOPOLOGY: linear
US-09-033-537A-1

```

Rest Local Similarity 47.1%; Pred. No. 8.1e-30;
Matches 77; Conservative 33; Mismatches 42; Indcls 6; Gaps 4;

Qy 2 SGGVKVOYKNNDSPGDNQIKPGLQLVNTGSSVDLSTVTVRYWFTRDGGSSTLVYNCDW 61
: : ||| ||| : ||| ||| : | : ||| : ||| ||| : ||| : ||
Db 402 TGNLVQYKVCDTSATDNQMKPSEINKNNGTTPVNI,SLCLKIKYYFTKD-CTADMSASFDM 460
: : ||| : | : ||| ||| : ||| ||| : ||| ||| : ||| ||| : ||| ||| : ||

Qy 62 AAMCGNIRASFSGVSNPATPTADTYLQLSFT--CGTLAAGSGTGEIQNRVKNKSDMSNFDE 119
: | : | : | : | : ||| ||| : ||| ||| : ||| ||| : ||| ||| : ||| ||| : ||
Db 461 AQIGASNVSAAF--ANFTGSDNTDYVELLSFAGSGSIPAGGTGDIQLRLMYTKTDSNFNE 518
: | : | : | : | : ||| ||| : ||| ||| : ||| ||| : ||| ||| : ||| ||| : ||

Qy 120 TNDYSY-CTNTAFODMTKVTYVYVNGRLVMGTPEP 151
||||| | ||| : ||| : ||| : ||| : ||| ||| |
Db 519 ANDYSYDGAKTAYADNRVTLHQNGTLVMGTTP 551
||||| | ||| : ||| : ||| : ||| : ||| ||| |

RESULT 2
US-07-862-588B-2
: Sequence 2, Application US/07862588B
: Patent No. 5916796
: GENERAL INFORMATION:
: APPLICANT: Joergensen, Per Linna
: APPLICANT: Sch Ieln, Martin
: APPLICANT: Hansen, Christian
: TITLE OF INVENTION: An Enzyme Exhibiting Cellulase Activity
: NUMBER OF SEQUENCES: 7
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: NO. 59167960 NO. 59167960 No. 59167960 disk of No. 5916796th America, Inc.
: STREET: 405 Lexington Avenue, 62nd floor
: CITY: New York
: STATE: New York
: COUNTRY: U.S.A.
: ZIP: 10017
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentin Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/07/862.588B
: FILING DATE: 19920727
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: DK 164/90
: FILING DATE: 19-JAN-1990
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: PCT/DK91/00013
: FILING DATE: 18-JAN-1991
: ATTORNEY/AGENT INFORMATION:
: NAME: Zelson, Steve T. / Lambiris, Elias J.
: REGISTRATION NUMBER: 30,335 / 33,728
: REFERENCE/DOCKET NUMBER: 3425.204-US
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 212 867 0123
: TELEFAX: 212 867 0298
: INFORMATION FOR SEQ ID NO: 2:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 700 amino acids
: TYPE: AMINO ACID
: TOPOLOGY: linear
: MOLECULE TYPE: protein
US-07-862-588B-2

Query Match 45.1%; Score 373.5; DB 2; Length 700;
Best Local Similarity 49.7%; Pred. No. 1.3e-29;
Matches 77; Conservative 33; Mismatches 39; Indcls 7; Gaps 5;

Qy 1 VSGVKVOYKNND-SAPCDNQIKPGLQLVNTGSSVDLSTVTVRYWFTRDGGSSTLVYNC 59
: : ||| ||| : ||| ||| : | : ||| : ||| ||| : ||| : ||
Db 549 VNSDLVYQYKDGDRNNATDNQIKPHEIQNRKGTSPVDLSLTLRYFTKD--SSAAMNGWI 607
: ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : |||

Qy 60 DWAMCCGNIRASFGSVNPATPTADTYLQLSFT--GGTLAAGSGTGEIQNRVKNKSDMSNF 117
: ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : |||

```


; PRIOR APPLICATION NUMBER: 60/067,249
; PRIOR FILING DATE: 1997-12-02
; PRIOR APPLICATION NUMBER: 60/067,240
; PRIOR FILING DATE: 1997-12-02
; PRIOR APPLICATION NUMBER: 09/073,684
; PRIOR FILING DATE: 1998-05-06
; PRIOR APPLICATION NUMBER: 09/184,217
; PRIOR FILING DATE: 1998-11-02
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12
; LENGTH: 493
; TYPE: PRT
; ORGANISM: Clostridium thermocellum
US-09-198-955A-12

Query Match 43.1% Score 357.5; DB 4; Length 493;
Best Local Similarity 42.3%; Pred. No. 3.3e-28;
Matches 69; Conservative 36; Mismatches 47; Indels 11; Gaps 3;
QY 1 VSGGVKVOYKNNDSAPGDNOIKPCLQIVNTGSSSDVLSITVTVRYWFTRDGSGSSTLVYNCD 60
DB 331 VSGNLKVEFYNSNPSTNTSINPQFKVTNTGSSAIDLSKLTLYYYTVDCOKDOTFW-CD 389
QY 61 WAAM-----CCGNIRASFGSVNPATPTADTYLQLSFTGGTTLAAGSGTGEIQNRVKN 111
DB 390 HAAIIGSNGSYNGITSNVKGTFVKMSSTNNADTYLFIPTGGTLEPCAHV-QIOGRFAK 448
QY 112 SDMSNFDENYSGTNTAFODMTKVTYVYVNGRLVWGTPEPSGT 154
DB 449 NDMSNTQSDNDYSFKRSQFVENDQVDTAYLNGVLVWCKEPPGS 491

RESULT 5
US-09-694-531-12
; Sequence 12, Application US/09694531
; Patent No. 6368843
; GENERAL INFORMATION:
; APPLICANT: Andersen, Lene N.
; APPLICANT: Schuelein, Martin
; APPLICANT: Lange, Niels Erik
; APPLICANT: Bjornvad, Mads E.
; APPLICANT: Moller, Soren
; APPLICANT: Glad, Sanne O. S.
; APPLICANT: Kauppinen, Markus S.
; APPLICANT: Schnorr, Kirk
; APPLICANT: Kongsbak, Lars
; TITLE OF INVENTION: No. 6368843el Pectate Lyases
; FILE REFERENCE: 5378.200-US
; CURRENT APPLICATION NUMBER: US/09/694,531
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 09/198,955
; PRIOR FILING DATE: 1998-11-24
; PRIOR APPLICATION NUMBER: 1343/97
; PRIOR FILING DATE: 1997-11-24
; PRIOR APPLICATION NUMBER: 1344/97
; PRIOR FILING DATE: 1997-11-24
; PRIOR APPLICATION NUMBER: 60/067,249
; PRIOR FILING DATE: 1997-12-02
; PRIOR APPLICATION NUMBER: 60/067,240
; PRIOR FILING DATE: 1997-12-02
; PRIOR APPLICATION NUMBER: 09/073,684
; PRIOR FILING DATE: 1998-05-06
; PRIOR APPLICATION NUMBER: 09/184,217
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12
; LENGTH: 493
; TYPE: PRT
; ORGANISM: Clostridium thermocellum
US-09-694-531-12

Query Match 43.1% Score 357.5; DB 4; Length 493;
Best Local Similarity 42.3%; Pred. No. 3.3e-28;
Matches 69; Conservative 36; Mismatches 47; Indels 11; Gaps 3;
QY 1 VSGGVKVOYKNNDSAPGDNOIKPCLQIVNTGSSSDVLSITVTVRYWFTRDGSGSSTLVYNCD 60
DB 331 VSGNLKVEFYNSNPSTNTSINPQFKVTNTGSSAIDLSKLTLYYYTVDCOKDOTFW-CD 389
QY 61 WAAM-----CCGNIRASFGSVNPATPTADTYLQLSFTGGTTLAAGSGTGEIQNRVKN 111
DB 390 HAAIIGSNGSYNGITSNVKGTFVKMSSTNNADTYLFIPTGGTLEPCAHV-QIOGRFAK 448
QY 112 SDMSNFDENYSGTNTAFODMTKVTYVYVNGRLVWGTPEPSGT 154
DB 449 NDMSNTQSDNDYSFKRSQFVENDQVDTAYLNGVLVWCKEPPGS 491

RESULT 6
US-09-670-141-10
; Sequence 10, Application US/09670141
; Patent No. 6429000
; GENERAL INFORMATION:
; APPLICANT: Andersen, Lene N.
; APPLICANT: Schuelein, Martin
; APPLICANT: Lange, Niels Erik K.
; APPLICANT: Bjornvad, Mads E.
; APPLICANT: Schnorr, Kirk
; TITLE OF INVENTION: Pectin Degrading Enzymes From Bacillus
; FILE REFERENCE: 5377.200-US
; CURRENT APPLICATION NUMBER: US/09/670,141
; CURRENT FILING DATE: 2000-09-26
; PRIOR FILING DATE: 1998-11-24
; PRIOR APPLICATION NUMBER: 09/198,956
; PRIOR FILING DATE: 1997-11-24
; PRIOR APPLICATION NUMBER: 60/067,240
; PRIOR FILING DATE: 1997-12-02
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 10
; LENGTH: 493
; TYPE: PRT
; ORGANISM: Bacillus licheniformis
US-09-670-141-10

Query Match 43.1% Score 357.5; DB 4; Length 493;
Best Local Similarity 42.3%; Pred. No. 3.3e-28;
Matches 69; Conservative 36; Mismatches 47; Indels 11; Gaps 3;
QY 1 VSGGVKVOYKNNDSAPGDNOIKPCLQIVNTGSSSDVLSITVTVRYWFTRDGSGSSTLVYNCD 60
DB 331 VSGNLKVEFYNSNPSTNTSINPQFKVTNTGSSAIDLSKLTLYYYTVDCOKDOTFW-CD 189
QY 61 WAAM-----CCGNIRASFGSVNPATPTADTYLQLSFTGGTTLAAGSGTGEIQNRVKN 111
DB 390 HAAIIGSNGSYNGITSNVKGTFVKMSSTNNADTYLFIPTGGTLEPCAHV-QIOGRFAK 448
QY 112 SDMSNFDENYSGTNTAFODMTKVTYVYVNGRLVWGTPEPSGT 154
DB 449 NDMSNTQSDNDYSFKRSQFVENDQVDTAYLNGVLVWCKEPPGS 491

RESULT 7
PCT-US95-13813-9
; Sequence 9, Application PC/TUS9513813
; GENERAL INFORMATION:
; APPLICANT: Yeda Research and Development Co. Ltd.
; APPLICANT: Ramot University Authority for Applied
; APPLICANT: Research and Industrial Development Ltd.
; APPLICANT: Technion Research and Development Foundation Ltd.
; APPLICANT: Bayer, Edward A.
; APPLICANT: Morag, Ely

APPLICANT: Wilchek, Meir
APPLICANT: Lamed, Raphael
APPLICANT: Shoham, Yuval
TITLE OF INVENTION: MODIFIED CELLULOSE-BINDING DOMAIN ((BD))
NUMBER OF SEQUENCES: 9
CORRESPONDENCE ADDRESS:
ADDRESSEE: Browdy and Neimark
CITY: Washington
STATE: D.C.
STREET: 419 Seventh Street N.W., Ste. 300
ZIP: 20004

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC Compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICANT NUMBER: PCT/US95/13813
FILING DATE:

CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Browdy, Roger L.
REGISTRATION NUMBER: 25,618
REFERENCE/DOCKET NUMBER: BAYER-3
TELEPHONE: (202) 628-5197
TELEFAX: (202) 737-3528
TELEX: 248633

INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 167 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
PCT-US95-13813-9

Query Match 43.0%; Score 356.5; DB 5; Length 167;
Best Local Similarity 42.3%; Pred. No. le-28;
Matches 69; Conservative 36; Mismatches 47; Indels 11; Gaps 3;

QY 1 VSGGVKVOYKNNDSAPGDNOIKPGLQVLNTGSSVDLSTVTVRYWFTRDGSGSTLYVNC 60
DB 5 VSGNLKVEFYNSPDTTNSINPOKVTNTGSSAIDLSKLTLLRYVYTVDGQKQDTFW-CD 63
QY 61 WAAM-----GCCNIRASFGSVNPATPTADTYLQLSFTGCTLAAGSGTGIEIONRVK 111
DB 64 HAAIIGNSGYNIGITSNVKGTVPYKMSSTNNAITYLEISFTGCTLEPGARV-QIOGRFAK 122

QY 112 SDMSNFDENDSYGTNTAFQDMTKVTYVYVNGRLVWGTPEPGT 154
DB 123 NDMSNYTQSDNYSFKXSQFVENDQVYATVINGVLVWCKEPGCS 165

RESULT 8
US-09-136-574A-47
Sequence 47, Application US/09136574A
Patent No. 6294366
GENERAL INFORMATION:
APPLICANT: Farrington, Graham K.
Anderson, Paige
Gibbs, Moreland
Bergquist, Peter
Daniels, Roy
Morgan, Hugh W.
Williams, Diane P.

TITLE OF INVENTION: Compositions and Methods for
Treating Cellulose Containing
Cellulase Enzyme Compositions
NUMBER OF SEQUENCES: 49
CORRESPONDENCE ADDRESS:
ADDRESSEE: Browdy and Neimark
CITY: Washington
STATE: D.C.
STREET: 419 Seventh Street N.W., Ste. 300
ZIP: 20004

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC Compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICANT NUMBER: PCT/US95/13813
FILING DATE:

CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Browdy, Roger L.
REGISTRATION NUMBER: 25,618
REFERENCE/DOCKET NUMBER: BAYER-3
TELEPHONE: (202) 628-5197
TELEFAX: (202) 737-3528
TELEX: 248633

INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 167 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
PCT-US95-13813-9

Query Match 43.0%; Score 356.5; DB 5; Length 167;
Best Local Similarity 42.3%; Pred. No. le-28;
Matches 69; Conservative 36; Mismatches 47; Indels 11; Gaps 3;

QY 1 VSGGVKVOYKNNDSAPGDNOIKPGLQVLNTGSSVDLSTVTVRYWFTRDGSGSTLYVNC 60
DB 5 VSGNLKVEFYNSPDTTNSINPOKVTNTGSSAIDLSKLTLLRYVYTVDGQKQDTFW-CD 63
QY 61 WAAM-----GCCNIRASFGSVNPATPTADTYLQLSFTGCTLAAGSGTGIEIONRVK 111
DB 64 HAAIIGNSGYNIGITSNVKGTVPYKMSSTNNAITYLEISFTGCTLEPGARV-QIOGRFAK 122

QY 112 SDMSNFDENDSYGTNTAFQDMTKVTYVYVNGRLVWGTPEPGT 154
DB 123 NDMSNYTQSDNYSFKXSQFVENDQVYATVINGVLVWCKEPGCS 165

RESULT 8
US-09-136-574A-47
Sequence 47, Application US/09136574A
Patent No. 6294366
GENERAL INFORMATION:
APPLICANT: Farrington, Graham K.
Anderson, Paige
Gibbs, Moreland
Bergquist, Peter
Daniels, Roy
Morgan, Hugh W.
Williams, Diane P.

TITLE OF INVENTION: Compositions and Methods for
Treating Cellulose Containing
Cellulase Enzyme Compositions
NUMBER OF SEQUENCES: 49
CORRESPONDENCE ADDRESS:
ADDRESSEE: Browdy and Neimark
CITY: Washington
STATE: D.C.
STREET: 419 Seventh Street N.W., Ste. 300
ZIP: 20004

ADDRESSEE: Howson and Howson
STREET: Spring House Corporate Center, P.O. Box 457
CITY: Spring House
STATE: PA
COUNTRY: USA
ZIP: 19477

COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
CURRENT APPLICATION DATA:
APPLICANT NUMBER: US/09/136,574A
FILING DATE: 19-Aug-1998
CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/932,571
FILING DATE: September 19, 1997
ATTORNEY/AGENT INFORMATION:
NAME: Bak, Mary E.

REGISTRATION NUMBER: 31,215
REFERENCE/DOCKET NUMBER: 1997US001/CIP
TELECOMMUNICATION INFORMATION:
TELEPHONE: 215-540-9200
TELEFAX: 215-540-5818
TELEX: <Unknown>

INFORMATION FOR SEQ ID NO: 47:
SEQUENCE CHARACTERISTICS:
LENGTH: 616 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 47:
US-09-136-574A-47

Query Match 42.1%; Score 349; DB 4; Length 616;
Best Local Similarity 43.8%; Pred. No. 3.2e-27;
Matches 71; Conservative 26; Mismatches 49; Indels 16; Gaps 4;

QY 1 VSGGVKVOYKNNDSAPGDNOIKPGLQVLNTGSSVDLSTVTVRYWFTRDGSGSTLYVNC 60
DB 1 MCGSGVLVYKNNETSASTGSIRPMFKIVNGSGSSVDLSRVKIRYWTVDGDKPQSAV-CD 54

QY 61 WAAMCGCNRASFGSVNPATPTADTYLQLSFTGCT--LAAGSGTGIEIONRVKSDMSNFD 118
DB 60 WAQIGASNTFTFVKLSGSGVSGADYILEVGFSSGACQLQPGKDTGDIQVRFNKNDMSNYN 119

QY 119 ETNDYS-----YGTNTAFQDMTKVTYVYVNGRLVWGTPEPG 153
DB 120 QADDWSWLOSMTNYGEN-----AKVTLYVDGVLVWGOEPG 155

RESULT 9
US-09-136-574A-44
Sequence 44, Application US/09136574A
Patent No. 6294366
GENERAL INFORMATION:
APPLICANT: Farrington, Graham K.
Anderson, Paige
Gibbs, Moreland
Bergquist, Peter
Daniels, Roy
Morgan, Hugh W.
Williams, Diane P.

TITLE OF INVENTION: Compositions and Methods for
Treating Cellulose Containing
Cellulase Enzyme Compositions
NUMBER OF SEQUENCES: 49
CORRESPONDENCE ADDRESS:
ADDRESSEE: Howson and Howson
STREET: Spring House Corporate Center, P.O. Box 457
CITY: Spring House
STATE: PA

```
; COUNTRY: USA
; ZIP: 19477
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/136,574A
; FILING DATE: 19-Aug-1998
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/932,571
; FILING DATE: September 19, 1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Bak, Mary E.
; REGISTRATION NUMBER: 31,215
; REFERENCE/DOCKET NUMBER: 1997US001/CIP
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 215-540-9200
; TELEFAX: 215-540-5818
; TELEX: <Unknown>
; INFORMATION FOR SEQ ID NO: 44:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1751 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 44:
US-09-136-574A-44

Query Match 42.0%; Score 348; DB 4; Length 1751;
Best Local Similarity 44.7%; Pred. No. 1.6e-26;
Matches 71; Conservative 25; Mismatches 47; Indels 16; Gaps 4;

Qy 4 GVKVOYKNDSAPGDNQIKPGQLQLVNTGSSVVDLSTVTVRYWFTRDGSGSSTLVYNCWAA 63
Db 678 GVKVLYKNNETSASTGSRPFKIVNGSSVVDLSRVKIRIYWTVDGDKPOSAY-CDWAA 736

Qy 64 MCGGNIRASFGSVNPATPTADTYLQLSFTGTT--LAAGSGTGEIQNRVKNKSDMSNFDFTN 121
Db 737 IGASNVTFNFVKLSGSGVADYILEVGFSSGAGLOPKDGTGDIQVRFNKNKDSNYSNQAD 796

Qy 122 DYS-----YGTNTAFQDWTKVTVYVNGRLVWGTEPSG 153
Db 797 DWSWLQSWMTNYGEN-----AKVTLYVDGVLVWGQEPGG 829

RESULT 10
US-09-136-574A-43
; Sequence 43, Application US/09136574A
; Patent No. 6294366
; GENERAL INFORMATION:
; APPLICANT: Farrington, Graham K.
; APPLICANT: Anderson, Paige
; APPLICANT: Gibbs, Moreland
; APPLICANT: Bergquist, Peter
; APPLICANT: Daniels, Roy
; APPLICANT: Morgan, Hugh W.
; APPLICANT: Williams, Diane P.
; TITLE OF INVENTION: Compositions and Methods for
; Treating Cellulose Containing Fabrics Using Truncated
; Cellulase Enzyme Compositions
; NUMBER OF SEQUENCES: 49
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Howson and Howson
; STREET: Spring House Corporate Center, P.O. Box 457
; CITY: Spring House
; STATE: PA
; COUNTRY: USA
; ZIP: 19477
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
```

```
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/136,574A
; FILING DATE: 19-Aug-1998
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/932,571
; FILING DATE: September 19, 1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Bak, Mary E.
; REGISTRATION NUMBER: 31,215
; REFERENCE/DOCKET NUMBER: 1997US001/CIP
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 215-540-9200
; TELEFAX: 215-540-5818
; TELEX: <Unknown>
; INFORMATION FOR SEQ ID NO: 43:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1426 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: No. 6294366e
; SEQUENCE DESCRIPTION: SEQ ID NO: 43:
US-09-136-574A-43

Query Match 41.0%; Score 340; DB 4; Length 1426;
Best Local Similarity 43.4%; Pred. No. 7.8e-26;
Matches 69; Conservative 26; Mismatches 48; Indels 16; Gaps 4;

Qy 4 GVKVOYKNDSAPGDNQIKPGQLQLVNTGSSVVDLSTVTVRYWFTRDGSGSSTLVYNCWAA 63
Db 413 GLKVLVKNNETSASTGSRPFKIVNGSSVVDLSRVKIRIYWTVDGDKPOSAY-CDWAA 471

Qy 64 MCGGNIRASFGSVNPATPTADTYLQLSFTGTT--LAAGSGTGEIQNRVKNKSDMSNFDFTN 121
Db 472 IGASNVTFNFVKLSGSGVADYILEVGFSSGAGLOPKDGTGDIQVRFNKNKDSNYSNQAD 511

Qy 122 DYS-----YGTNTAFQDWTKVTVYVNGRLVWGTEPSG 153
Db 532 DWSWLQSWMTNYGEN-----AKVTLYVDGVLVWGQEPGG 564

RESULT 11
US-08-048-164A-2
; Sequence 2, Application US/08048164A
; Patent No. 5496934
; GENERAL INFORMATION:
; APPLICANT: Shoseyov, Oded
; APPLICANT: Shplegl, Itai
; APPLICANT: Goldstein, Marc A.
; APPLICANT: Doi, Roy H.
; TITLE OF INVENTION: CELLULOSE BINDING DOMAIN
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PENNIE & EDMONDS
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/048,164A
; FILING DATE: 14-APR-1993
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Mistrock, S. Leslie
```

```

: REGISTRATION NUMBER: 18,872
: REFERENCE/DOCKET NUMBER: 7809-003
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (212) 790-9090
: TELEFAX: (212) 869-8864/9741
: TELEX: 66141 PENNIE
: INFORMATION FOR SEQ ID NO: 2:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 162 amino acids
: TYPE: amino acid
: TOPOLOGY: linear
: MOLECULE TYPE: protein
: US-08-048-164A-2

Query Match          27.38; Score 226; DB 1; Length 162;
Best Local Similarity 34.88; Pred. No. 1.5e-15;
Matches 57; Conservative 31; Mismatches 58; Indels 18; Gaps 7;

QY 2 SGGVKVQYKKNDSAPGDNIQKPGIQLVNTGSSSVDLSLTVTVYWFTRDGSSTLVYNCDW 61
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 3 TSSMSVEFYNSKSAQTNSITPIIKITNTSPDLNLNDVKVRYVYTSQTGQTQTFW-CDH 61
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 62 AAMCGGN-----IRASF--GSVNPATPTADYIQLSFTGCTLAAGSGTCE---IONK 108
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 62 AGALLGNSYVDNTSKVTANFVKETASP-TSTYDTYVEFGFASG--AATLKKGFITIQGR 118
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 109 VNKSDWSNFDICNDYSYGTNTAFQDWT-KVTVYVNGRLVWGCTEP 151
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 119 ITKSDWSNYTOTNDYSFSDASSTPVNPVKVTGYICGAKVLTGAP 162
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 12
US-08-460-462-2
: Sequence 2, Application US/08460462
: Patent No. 5670623
: GENERAL INFORMATION:
: APPLICANT: Shoseyov, Oded
: APPLICANT: Shpiegl, Itai
: APPLICANT: Goldstein, Marc A.
: APPLICANT: Doi, Roy H.
: TITLE OF INVENTION: METHODS OF USE OF CELLULOSE BINDING DOMAIN PROTEINS
: NUMBER OF SEQUENCES: 21
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: PENNIE & EDMONDS
: STREET: 1155 Avenue of the Americas
: CITY: New York
: STATE: New York
: COUNTRY: U.S.A.
: ZIP: 10036
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patent In Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/460.462
: FILING DATE: concurrently herewith
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/048,164
: FILING DATE: 14-APR-1993
: ATTORNEY/AGENT INFORMATION:
: NAME: Mlsrock, S. Leslie
: REGISTRATION NUMBER: 18,872
: REFERENCE/DOCKET NUMBER: 7809-006
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (212) 790-9090
: TELEFAX: (212) 869-8864/9741
: TELEX: 66141 PENNIE
: INFORMATION FOR SEQ ID NO: 2:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 162 amino acids
: TYPE: amino acid

```

[illegible]

QY 2 SGGVKVQYKNNDSAPGDNQIKPGLQLVNTGSSVDLSTVTYVWFTRDGGSTLVYNCDW 61

Db	3	TSSMSVEFYNSNSKAQTNSTPILIKITNTSDSLNLDNVKRYYYTSDGTQGTGF-CDII	61
QY	62	AAMCCGN-----IRASH--GSVNPATPATYTYQLSFTGCTLAAGSTGE---IQNR	108
		: : : : : : : : : : : : : : :	
Db	62	AGALLGNSYVDNTSKVTANFKVETASP-TSTYDTYVEFGFASG--AATLUKQOFTIQGR	118
QY	109	VNKSQMSNDETDNDYSYGTNTAFQDWT-KVTYVVGRLVMGTGP	151
Db	119	ITKSDMSNYTQTDNDYGFDAASSTPPVNPVKVTGYIGCAKVLGTAP	162
		: : : : : : : : : : :	

RESULT 14
US-08-460-458-2
: Sequence 2, Application US/08460458
: Patent No. 5738984
: GENERAL INFORMATION:
: APPLICANT: Shoscyov, Oded
: TITLE OF INVENTION: KITS AND METHODS OF DETECTION USING CELLULOSE BINDING DOMAIN
: NUMBER OF SEQUENCES: 21
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: PENNIE & EDMONDS
: STREET: 1155 Avenue of the Americas
: CITY: New York
: STATE: New York
: COUNTRY: U.S.A.
: ZIP: 10036
: COMPUTER READABLE FORM:

[illegible]

RESULT 15
US-08-460-455-2

```

: Sequence 2, Application US/08460455
: Patent No. 5837814
: GENERAL INFORMATION:
: APPLICANT: Shoseyov, Oded
: APPLICANT: Shoseyov, Itai
: APPLICANT: Goldstein, Marc A.
: APPLICANT: Dol, Roy H.
: TITLE OF INVENTION: CELLULOSE BINDING DOMAIN PROTEINS
: NUMBER OF SEQUENCES: 21
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: PENNIE & EDMONDS
: STREET: 1155 Avenue of the Americas
: CITY: New York
: STATE: New York
: COUNTRY: U.S.A.
: ZIP: 10036
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/460,455
: FILING DATE: concurrently herewith
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/048,164
: FILING DATE: 14-APR-1993
: ATTORNEY/AGENT INFORMATION:
: NAME: Misrock, S. Leslie
: REGISTRATION NUMBER: 18,872
: REFERENCE/DOCKET NUMBER: 7809-009
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (212) 790-9090
: TELEFAX: (212) 869-8864/9741
: TELEX: 66141 PENNIE
: INFORMATION FOR SEQ ID NO: 2:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 162 amino acids
: TYPE: amino acid
: TOPOLOGY: linear
: MOLECULE TYPE: protein
: US-08-460-455-2

```

```

Query Match      27.3%   Score 226;   DB 2:   Length 162;
Best Local Similarity 34.8%;   Pred. No. 1.5e-15;
Matches 57; Conservative 31; Mismatches 58; Indels 18; Gaps 17;

Qy      2  SGGVKVQYKNDSPAGDNOIKPGLQLQVNTGSSSDVLSTVTVRYVWTRDGGSSSTLVYNCDW 61
      :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db      3  TSSMSVEFYNSKNQAQNSITPIIKITNTSDSLNLNDVKVRYVYTSCTOGQTFW-CDH 61
      :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :

Qy      62  AAMCCGN-----IRASF--GSVNPAATPADTYQLQSLFTGGTLAAGGSGTGE---IQNR 108
      :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db      62  AGALLGNSYVDNTSKVFANFVKETASP-TSTDYTYVVFEGFASG--AATLKKGFQFITIQGR 118
      :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :

Qy      109  VNKSDWSNFDETNDYSYGTNTAFODWT-KVTVYVNGRLVWGCTEP 151
      :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db      119  ITKSDWSNYTOTNDYSFDASSPTVPVNPVKVTCYIGGAKVLGCTP 162
      :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :

Search completed: November 13, 2002, 11:56:06
Job time : 5.75718 secs

```


GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 13, 2002, 11:55:11 : Search time 2.98707 Seconds
(without alignments)
776.473 Million cell updates/sec

Title: US-09-917-378-4

Perfect score: 829

Sequence: 1 VSGVKVQYKKNDSAPCDNQ.....TKVTYVNGRLVWGTPSGT 154

Scoring table: BLOSUM62

Gapop 10.0 , Capext 0.5

Searched: 97044 seqs, 15060890 residues

Total number of hits satisfying chosen parameters: 97044

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published_Applications_AA:*
1: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep:*
2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep:*
3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep:*
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep:*
5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep:*
6: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep:*
7: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep:*
8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep:*
9: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep:*
10: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep:*
11: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep:*
12: /cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB.pep:*
13: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep:*
14: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	357.5	43.1	493	12	US-10-072-152-12
2	90.5	10.9	986	10	US-09-747-835A-53
3	90.5	10.9	1325	10	US-09-747-835A-24
4	90.5	10.9	1346	10	US-09-747-835A-20
5	90.5	10.9	1346	10	US-09-747-835A-61
6	90.5	10.9	1371	10	US-09-747-835A-25
7	80	9.7	1599	9	US-10-092-880-9
8	80	9.7	1600	9	US-10-092-880-10
9	75.5	9.1	1349	10	US-09-747-835A-52
10	75.5	9.1	2353	10	US-09-797-862-33
11	74.5	9.0	871	10	US-09-886-488-21
12	72.5	8.7	69	10	US-09-764-860-377
13	72.5	8.7	1333	10	US-09-815-242-10936
14	70.5	8.5	334	10	US-09-870-521-4
15	70	8.4	300	10	US-09-815-242-11359
16	70	8.4	300	10	US-09-881-752A-202
17	70	8.4	1974	9	US-09-895-913A-12
18	69	8.3	294	9	US-09-981-353-108
19	68.5	8.3	185	9	US-09-970-616-2

Query Match 43.1%; Score 357.5; DB 12; Length 493;
Best Local Similarity 42.3%; Pred. No. 4.5e-30;

ALIGNMENTS

RESULT 1

US-10-072-152-12

: Sequence 12, Application US/10072152

: Patent No. US20020142438A1

: GENERAL INFORMATION:

: APPLICANT: Andersen, Lene N.

: APPLICANT: Schulein, Martin

: APPLICANT: Lange, Niels E.

: APPLICANT: Bjornvad, Mads E.

: APPLICANT: Moller, Soren

: APPLICANT: Glad, Sanne O. S.

: APPLICANT: Kauppinen, Markus S.

: APPLICANT: Schmor, Kirk

: APPLICANT: Kongsbak, Lars

: TITLE OF INVENTION: No. US20020142438A1e1 Pectate Lyases

: FILE REFERENCE: 5378.200-US

: CURRENT APPLICATION NUMBER: US/10/072.152

: PRIOR FILING DATE: 2002-02-07

: PRIOR APPLICATION NUMBER: US/09/198.955

: PRIOR FILING DATE: 1998-11-24

: PRIOR APPLICATION NUMBER: 1343/97

: PRIOR FILING DATE: 1997-11-24

: PRIOR APPLICATION NUMBER: 1344/97

: PRIOR FILING DATE: 1997-11-24

: PRIOR APPLICATION NUMBER: 60/067.249

: PRIOR FILING DATE: 1997-12-02

: PRIOR APPLICATION NUMBER: 60/067.240

: PRIOR FILING DATE: 1997-12-02

: PRIOR APPLICATION NUMBER: 09/073.684

: PRIOR FILING DATE: 1998-05-06

: PRIOR APPLICATION NUMBER: 09/184.217

: PRIOR FILING DATE: 1998-11-02

: NUMBER OF SEQ ID NOS: 32

: SOFTWARE: FastSeq for Windows Version 4.0

: SEQ ID NO 12

: LENGTH: 493

: TYPE: PRT

: ORGANISM: Clostridium thermocellum

US-10-072-152-12

```

Matches 69: Conservative 36: Mismatches 47: Indels 11: Gaps 3:
Oy 1 VSGGVKVOYKKNDSAPGDNIQKPGQLVNTGSSSVDLSTVTIVRWFTGRDGSSTLVYNCD 60
   ||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db 331 VSGNLKVEFYNSNPSDDTNSINPOKFKVTNTGSSAIDLKSLTLRYYYYVDGQKQDTFW-CD 389

Oy 61 WAAM-----CCGNIRASEFGSVNPATPDYHQLSLFTGGTLAAGSGTGETQNRVVK 111
   ||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db 390 HAAIIGSGSYNGITSNVKYGTFVKMSSSTNADTYLEISFTGGTLEPGAHV-QIOGREAK 448

Oy 112 SDWSNFDETNDYSYCTNTAFQDWTKTVTYVNGRLVWGTPEPSC 154
   ||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db 449 NDWSNYTOSNDYSPKRSQFVENDQVATYLVGLVWGREPGGS 491

RESULT 2
US-09-747-835A-53
: Sequence 53, Application US/09747835A
: Patent No. US20020146692A1
: GENERAL INFORMATION:
: APPLICANT: Yamazaki, Victoria
: APPLICANT: Tang, Y. Tom
: APPLICANT: Liu, Chenghua
: APPLICANT: Zhou, Ping
: APPLICANT: Wang, Dunrui
: APPLICANT: Zhang, Jie
: APPLICANT: Ren, Feiyan
: APPLICANT: Asundi, Vinod
: APPLICANT: Drmanac, Radoje T
: TITLE OF INVENTION: METHODS AND MATERIALS RELATING TO G PROTEIN-COUPLED RECEPTORS
: FILE REFERENCE: HVS-37C1P
: CURRENT APPLICATION NUMBER: US/09/747,835A
: CURRENT FILING DATE: 2002-03-08
: PRIOR APPLICATION NUMBER: US 09/729,739
: PRIOR FILING DATE: 2000-12-04
: PRIOR APPLICATION NUMBER: US 09/653,450
: PRIOR FILING DATE: 2000-08-31
: PRIOR APPLICATION NUMBER: US 09/620,312
: PRIOR FILING DATE: 2000-07-19
: PRIOR APPLICATION NUMBER: US 09/598,042
: PRIOR FILING DATE: 2000-06-20
: PRIOR APPLICATION NUMBER: US 09/552,317
: PRIOR FILING DATE: 2000-04-25
: PRIOR APPLICATION NUMBER: US 09/488,725
: PRIOR FILING DATE: 2000-01-21
: NUMBER OF SEQ ID NOS: 63
: SOFTWARE: PatentIn version 3.0
: SEQ ID NO 53
: LENGTH: 986
: TYPE: prt
: ORGANISM: Homo sapiens
US-09-747-835A-53

```

```
Query Match      10.9% Score 90.5; DB 10; Length 986;
Best Local Similarity 29.2%; Pred. No. 0.16;
Matches 38; Conservative 15; Mismatches 42; Indels 35; Gaps 7;

QY    31  GSSVDLSTVTVRYWFTRDG-----GSS--TLVVNCD----WAAMGCCNIPASFGSNPA 79
       | : : : | | | | | | | | | | | | | | | | | | | | | | | |
Db     54  GTPTDIDSCSRVTLKADGTCTCPSGSGTIVITTCFISAYGARGSANIKVTFISVANL 113

QY    80  TPTADTVQLSFTGGTLAAGSTGEIONRVNK--SDMSNFDETDNDYSYGNTA----FQD 133
       | | | | | | | | | | | | | | | | | | | | | | | |
Db    114  TITPDPI-----SVSECONFSIKCISDSVSNDEV---YWNTSAGIKIYQR 155

QY    134  WTQKVTVVYNG 143
        : | : |
Db    156  FYTRTRYLDG 165
```

RESULT 3
US-09-747-835A-24

```
; Sequence 24, Application US/09747835A
; Patent No. US20020146692A1
; GENERAL INFORMATION:
; APPLICANT: Yamazaki, Victoria
; APPLICANT: Tang, Y. Tom
; APPLICANT: Liu, Chenghua
; APPLICANT: Zhou, Ping
; APPLICANT: Wang, Dunrui
; APPLICANT: Zhang, Jie
; APPLICANT: Ren, Feliyan
; APPLICANT: Asundi, Vinod
; APPLICANT: Dmanac, Radoje T
; TITLE OF INVENTION: METHODS AND MATERIALS RELATING TO G PROTEIN-COUPLED RE
; FILE OF INVENTION: LIKE) POLYPEPTIDES AND POLYNUCLEOTIDES
; FILE REFERENCE: HVS-37CIP
; CURRENT APPLICATION NUMBER: US/09/747,835A
; PRIOR FILING DATE: 2002-03-08
; PRIOR APPLICATION NUMBER: US 09/729,739
; PRIOR FILING DATE: 2000-12-04
; PRIOR APPLICATION NUMBER: US 09/653,450
; PRIOR FILING DATE: 2000-08-31
; PRIOR APPLICATION NUMBER: US 09/620,312
; PRIOR FILING DATE: 2000-07-19
; PRIOR APPLICATION NUMBER: US 09/598,042
; PRIOR FILING DATE: 2000-06-20
; PRIOR APPLICATION NUMBER: US 09/552,317
; PRIOR FILING DATE: 2000-04-25
; PRIOR APPLICATION NUMBER: US 09/488,725
; PRIOR FILING DATE: 2000-01-21
; NUMBER OF SEQ ID NOS: 63
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 24
; LENGTH: 1325
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-747-835A-24

Query Match          10.9%; Score 90.5; DB 10; Length 1325:
Best Local Similarity 29.2%; Pred.No. 0.23;
Matches 38; Conservative 15; Mismatches 42; Indels 35; Gaps

Qy      31  GSSSVDLSTVTVRYWFETRDG-----GSS--TLVVNCD---WAAMCCGNIRASFGSNVA 79
         |:|:::|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
Db       393  GPETDIDSSCSRTLKADGTQPSSGGTIVTYCEFIISAYGARGSANIKVTFISVANL 452
         ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

Qy      80  TPTADTYLQLSFETGGTLAAGSGTGTEIGNRVNK--SDWSNFDETNDYSYGNTA----FOD 133
         ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db       453  TLTPDPI-----SYSEGQNFSIKISDSVNDEV----YWNTSACIKIYOK 494
         ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

Qy      134  WTKVTYYVNG 143
         : | | |
Db       495  FYTRRYLDG 504
```

```

RESULT 4
US-09-747-835A-20
: Sequence 20. Application US/09747835A
: Patent No. US20020145692A1
: GENERAL INFORMATION:
: APPLICANT: Yamazaki, Victoria
: APPLICANT: Tang, Y. Tom
: APPLICANT: Liu, Chenghua
: APPLICANT: Zhou, Ping
: APPLICANT: Wang, Dunrui
: APPLICANT: Zhang, Jie
: APPLICANT: Ren, Feiyao
: APPLICANT: Asundi, Vinod
: APPLICANT: Dmanac, Radoje T
: TITLE OF INVENTION: METHODS AND MATERIALS RELATING TO G PROTEIN-COUPLED
: TITLE OF INVENTION: LIKE) POLYPEPTIDES AND POLYNUCLEOTIDES
: FILE REFERENCE: HYS-37CIP
: CURRENT APPLICATION NUMBER: US/09/747,835A
: CURRENT FILING DATE: 2002-03-08

```



```
; PRIOR APPLICATION NUMBER: US 09/729,739
; PRIOR FILING DATE: 2000-12-04
; PRIOR APPLICATION NUMBER: US 09/653,450
; PRIOR FILING DATE: 2000-08-31
; PRIOR APPLICATION NUMBER: US 09/620,312
; PRIOR FILING DATE: 2000-07-19
; PRIOR APPLICATION NUMBER: US 09/598,042
; PRIOR FILING DATE: 2000-06-20
; PRIOR APPLICATION NUMBER: US 09/552,317
; PRIOR FILING DATE: 2000-04-25
; PRIOR APPLICATION NUMBER: US 09/488,725
; PRIOR FILING DATE: 2000-01-21
; NUMBER OF SEQ ID NOS: 63
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 20
; LENGTH: 1346
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-747-835A-20

Query Match      10.9%  Score 90.5;  DB 10;  Length 1346;
Best Local Similarity 29.2%;  Pred. No. 0.23;
Matches 38;  Conservative 15;  Mismatches 42;  Indels 35;  Gaps 7;

Oy  31  GSSVDLSTVTYVRYWFTRDG-----GSS--TLVYNCD-----WAAMCGGNIRASFGSVNPA 79
      | : : : | | | | | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db  414  GTPETDIDSSCSRYTLKADGTCPCSGSGTTVIYTCEFIISAYGARGSANIKVTFISVANL 473

Oy  80  TPTADTYLQLSFTGCTLAAGGSGTEIQNRVKNK--SDWSNFDNDYSYGTNTA----FQD 133
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db  474  TITPDPi-----SVSEQNFSIKCISDVSNYDEV----YWNTSAGIKIYOR 515

Oy  134  WTKVTYVYVNG 143
      : | : |
Db  516  FYTTRRYLDG 525

RESULT 5
US-09-747-835A-61
; Sequence 61, Application US/09747835A
; Patent No. US20020146692A1
; GENERAL INFORMATION:
; APPLICANT: Yamazaki, Victoria
; APPLICANT: Tang, Y. Tom
; APPLICANT: Liu, Chonghua
; APPLICANT: Zhou, Ping
; APPLICANT: Wang, Dunrui
; APPLICANT: Zhang, Jie
; APPLICANT: Ren, Feiyan
; APPLICANT: Asundi, Vinod
; APPLICANT: Drmanac, Radoje T
; TITLE OF INVENTION: METHODS AND MATERIALS RELATING TO G PROTEIN-COUPLED RECEPTOR-LIKE
; FILE REFERENCE: HYS-37CIP
; CURRENT APPLICATION NUMBER: US/09/747,835A
; CURRENT FILING DATE: 2002-03-08
; PRIOR APPLICATION NUMBER: US 09/729,739
; PRIOR FILING DATE: 2000-12-04
; PRIOR APPLICATION NUMBER: US 09/653,450
; PRIOR FILING DATE: 2000-08-31
; PRIOR APPLICATION NUMBER: US 09/620,312
; PRIOR FILING DATE: 2000-07-19
; PRIOR APPLICATION NUMBER: US 09/598,042
; PRIOR FILING DATE: 2000-06-20
; PRIOR APPLICATION NUMBER: US 09/552,317
; PRIOR FILING DATE: 2000-04-25
; PRIOR APPLICATION NUMBER: US 09/488,725
; PRIOR FILING DATE: 2000-01-21
; NUMBER OF SEQ ID NOS: 63
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 61
; LENGTH: 1346
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-747-835A-25

Query Match      10.9%  Score 90.5;  DB 10;  Length 1371;
Best Local Similarity 29.2%;  Pred. No. 0.24;
Matches 38;  Conservative 15;  Mismatches 42;  Indels 35;  Gaps 7;

Oy  31  GSSVDLSTVTYVRYWFTRDG-----GSS--TLVYNCD-----WAAMCGGNIRASFGSVNPA 79
      | : : : | | | | | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db  438  GTPETDIDSSCSRYTLKADGTCPCSGSGTTVIYTCEFIISAYGARGSANIKVTFISVANL 497

Oy  80  TPTADTYLQLSFTGCTLAAGGSGTEIQNRVKNK--SDWSNFDNDYSYGTNTA----FQD 133
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db  498  TITPDPi-----SVSEQNFSIKCISDVSNYDEV----YWNTSAGIKIYOR 539

Oy  134  WTKVTYVYVNG 143
      : | : |
```

```
; ORGANISM: Homo sapiens
US-09-747-835A-61

Query Match      10.9%  Score 90.5;  DB 10;  Length 1346;
Best Local Similarity 29.2%;  Pred. No. 0.23;
Matches 38;  Conservative 15;  Mismatches 42;  Indels 35;  Gaps 7;

Oy  31  GSSVDLSTVTYVRYWFTRDG-----GSS--TLVYNCD-----WAAMCGGNIRASFGSVNPA 79
      | : : : | | | | | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db  414  GTPETDIDSSCSRYTLKADGTCPCSGSGTTVIYTCEFIISAYGARGSANIKVTFISVANL 473

Oy  80  TPTADTYLQLSFTGCTLAAGGSGTEIQNRVKNK--SDWSNFDNDYSYGTNTA----FQD 133
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db  474  TITPDPi-----SVSEQNFSIKCISDVSNYDEV----YWNTSAGIKIYOR 515

Oy  134  WTKVTYVYVNG 143
      : | : |
Db  516  FYTTRRYLDG 525

RESULT 6
US-09-747-835A-25
; Sequence 25, Application US/09747835A
; Patent No. US20020146692A1
; GENERAL INFORMATION:
; APPLICANT: Yamazaki, Victoria
; APPLICANT: Tang, Y. Tom
; APPLICANT: Liu, Chonghua
; APPLICANT: Zhou, Ping
; APPLICANT: Wang, Dunrui
; APPLICANT: Zhang, Jie
; APPLICANT: Ren, Feiyan
; APPLICANT: Asundi, Vinod
; APPLICANT: Drmanac, Radoje T
; TITLE OF INVENTION: METHODS AND MATERIALS RELATING TO G PROTEIN-COUPLED RECEPTOR-LIKE
; FILE REFERENCE: HYS-37CIP
; CURRENT APPLICATION NUMBER: US/09/747,835A
; CURRENT FILING DATE: 2002-03-08
; PRIOR APPLICATION NUMBER: US 09/729,739
; PRIOR FILING DATE: 2000-12-04
; PRIOR APPLICATION NUMBER: US 09/653,450
; PRIOR FILING DATE: 2000-08-31
; PRIOR APPLICATION NUMBER: US 09/620,312
; PRIOR FILING DATE: 2000-07-19
; PRIOR APPLICATION NUMBER: US 09/598,042
; PRIOR FILING DATE: 2000-06-20
; PRIOR APPLICATION NUMBER: US 09/552,317
; PRIOR FILING DATE: 2000-04-25
; PRIOR APPLICATION NUMBER: US 09/488,725
; PRIOR FILING DATE: 2000-01-21
; NUMBER OF SEQ ID NOS: 63
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 25
; LENGTH: 1371
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-747-835A-25

Query Match      10.9%  Score 90.5;  DB 10;  Length 1371;
Best Local Similarity 29.2%;  Pred. No. 0.24;
Matches 38;  Conservative 15;  Mismatches 42;  Indels 35;  Gaps 7;

Oy  31  GSSVDLSTVTYVRYWFTRDG-----GSS--TLVYNCD-----WAAMCGGNIRASFGSVNPA 79
      | : : : | | | | | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db  438  GTPETDIDSSCSRYTLKADGTCPCSGSGTTVIYTCEFIISAYGARGSANIKVTFISVANL 497

Oy  80  TPTADTYLQLSFTGCTLAAGGSGTEIQNRVKNK--SDWSNFDNDYSYGTNTA----FQD 133
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db  498  TITPDPi-----SVSEQNFSIKCISDVSNYDEV----YWNTSAGIKIYOR 539

Oy  134  WTKVTYVYVNG 143
      : | : |
```

Db 540 FYTTRYLDG 549

RESULT 7

US-10-092-880-9

: Sequence 9, Application US/10092880

: Patent No. US20020164354A1

: GENERAL INFORMATION:

: APPLICANT: Barenkamp, Stephen J.

: TITLE OF INVENTION: HIGH MOLECULAR WEIGHT SURFACE PROTEINS OF NON-TYPEABLE

: FILE REFERENCE: HAEMOPHILUS

: CURRENT APPLICATION NUMBER: US/10/092,880

: CURRENT FILING DATE: 2002-03-08

: PRIOR APPLICATION NUMBER: 09/155,614

: PRIOR FILING DATE: 1998-09-30

: PRIOR APPLICATION NUMBER: 08/617,697

: PRIOR FILING DATE: 1996-04-01

: PRIOR APPLICATION NUMBER: PCT/US97/04707

: PRIOR FILING DATE: 1997-04-01

: NUMBER OF SEQ ID NOS: 11

: SOFTWARE: PatentIn Ver. 2.1

: SEQ ID NO 9

: LENGTH: 1599

: TYPE: PRT

: ORGANISM: Haemophilus influenzae

US-10-092-880-9

Query Match

Best Local Similarity 9.7%; Score 80; DB 9; Length 1599;

Matches 37; Conservative 26; Mismatches 80; Indels 20; Gaps 5;

Qy 2

SGGVKVOYKKNDSAPGDNQIKPGIQLVNTGS-----SSVDLSTVTYRYWFTTRGGS-STLV 56

Db 1198 SGTVNISTKTGDIKGGIESTSCNVNITASGNTLKVSNTITGQDVT-----TADAGALTTTA 1253

Qy 57

YNCDAAMCGCINIRASFGSNPATPTADTYLQLSFTGCTLAAGSGTGEI-----QNRVN 110

Db 1254 GSTISATTGNANITTKTGDIKGVESSSGVTLVATGATLAVGNISGNTVTITADSGKLT 1313

Qy 111

KSDWSNFDETNDYSGTNTAFQDMTKVTYVYNGRLVWCTEPSG 153

Db 1314 STVGSTINGTNSVTSSQSGDIEGT-----ISGNTVNVNTASTG 1351

RESULT 8

US-10-092-880-10

: Sequence 10, Application US/10092880

: Patent No. US20020164354A1

: GENERAL INFORMATION:

: APPLICANT: Barenkamp, Stephen J.

: TITLE OF INVENTION: HIGH MOLECULAR WEIGHT SURFACE PROTEINS OF NON-TYPEABLE

: FILE REFERENCE: HAEMOPHILUS

: CURRENT APPLICATION NUMBER: US/10/092,880

: CURRENT FILING DATE: 2002-03-08

: PRIOR APPLICATION NUMBER: 09/155,614

: PRIOR FILING DATE: 1998-09-30

: PRIOR APPLICATION NUMBER: 08/617,697

: PRIOR FILING DATE: 1996-04-01

: PRIOR APPLICATION NUMBER: PCT/US97/04707

: PRIOR FILING DATE: 1997-04-01

: NUMBER OF SEQ ID NOS: 11

: SOFTWARE: PatentIn Ver. 2.1

: SEQ ID NO 10

: LENGTH: 1600

: TYPE: PRT

: ORGANISM: Haemophilus influenzae

US-10-092-880-10

Query Match

Best Local Similarity 9.7%; Score 80; DB 9; Length 1600;

Matches 37; Conservative 26; Mismatches 80; Indels 20; Gaps 5;

Qy 2

SGGVKVOYKKNDSAPGDNQIKPGIQLVNTGS-----SSVDLSTVTYRYWFTTRGGS-STLV 56

Db 1198 SGTVNISTKTGDIKGGIESTSCNVNITASGNTLKVSNTITGQDVT-----TADAGALTTTA 1253

Qy 57

YNCDAAMCGCINIRASFGSNPATPTADTYLQLSFTGCTLAAGSGTGEI-----QNRVN 110

Db 1254 GSTISATTGNANITTKTGDIKGVESSSGVTLVATGATLAVGNISGNTVTITADSGKLT 1313

Qy 111

KSDWSNFDETNDYSGTNTAFQDMTKVTYVYNGRLVWCTEPSG 153

Db 1314 STVGSTINGTNSVTSSQSGDIEGT-----ISGNTVNVNTASTG 1351

Qy 2 SGGVKVOYKKNDSAPGDNQIKPGIQLVNTGS-----SSVDLSTVTYRYWFTTRGGS-STLV 56

Db 1198 SGTVNISTKTGDIKGGIESTSCNVNITASGNTLKVSNTITGQDVT-----TADAGALTTTA 1254

Qy 57 YNCDAAMCGCINIRASFGSNPATPTADTYLQLSFTGCTLAAGSGTGEI-----QNRVN 110

Db 1254 GSTISATTGNANITTKTGDIKGVESSSGVTLVATGATLAVGNISGNTVTITADSGKLT 1314

Qy 111 KSDWSNFDETNDYSGTNTAFQDMTKVTYVYNGRLVWCTEPSG 153

Db 1314 STVGSTINGTNSVTSSQSGDIEGT-----ISGNTVNVNTASTG 1352

RESULT 9

US-09-747-835A-52

: Sequence 52, Application US/09747835A

: Patent No. US20020146692A1

: GENERAL INFORMATION:

: APPLICANT: Yamazaki, Victoria

: APPLICANT: Tang, Y. Tom

: APPLICANT: Liu, Chenchua

: APPLICANT: Zhou, Ping

: APPLICANT: Wang, Dunrui

: APPLICANT: Zhang, Jie

: APPLICANT: Ren, Feiyan

: APPLICANT: Asundi, Vinod

: APPLICANT: Drmanac, Radoje T

: TITLE OF INVENTION: METHODS AND MATERIALS RELATING TO G PROTEIN-COUPLED RECEPTOR-L

: FILE REFERENCE: HVS-37CIP

: CURRENT APPLICATION NUMBER: US/09/747,835A

: CURRENT FILING DATE: 2002-03-08

: PRIOR APPLICATION NUMBER: US 09/729,739

: PRIOR FILING DATE: 2000-12-04

: PRIOR APPLICATION NUMBER: US 09/653,450

: PRIOR FILING DATE: 2000-08-31

: PRIOR APPLICATION NUMBER: US 09/620,312

: PRIOR FILING DATE: 2000-07-19

: PRIOR APPLICATION NUMBER: US 09/598,042

: PRIOR FILING DATE: 2000-06-20

: PRIOR APPLICATION NUMBER: US 09/552,317

: PRIOR FILING DATE: 2000-04-25

: PRIOR APPLICATION NUMBER: US 09/488,725

: PRIOR FILING DATE: 2000-01-21

: NUMBER OF SEQ ID NOS: 63

: SOFTWARE: PatentIn version 3.0

: SEQ ID NO 52

: LENGTH: 1349

: TYPE: PRT

: ORGANISM: Rattus norvegicus

US-09-747-835A-52

Query Match

Best Local Similarity 9.1%; Score 75.5; DB 10; Length 1349;

Matches 30; Conservative 10; Mismatches 37; Indels 23; Gaps 4;

Qy 31

GSSSDLSSTVTYRYWFTTRDG-----GSS--TLVYNCD-----WAAMCGCINIRASFGSNPA 79

Db 412 GTPETDLESSCSTYTLKADGTQCPGSSGTTVITCEPVSVYGAKGSKNIAVTFTSVANL 471

Qy 80

TPTADTYLQLSFTGCTLAAGSGTGEIONRVNKSQSNFDE 119

Db 472 TITPD-----PISVSGQSFSITCLSDVSSPDE 499

RESULT 10

US-09-797-862-33

: Sequence 33, Application US/09797862

: Patent No. US20020102276A1

: GENERAL INFORMATION:

: APPLICANT: PEAK, IAN RICHARD ANSELM

: APPLICANT: JENNINGS, MICHAEL PAUL

```

; PRIOR FILING DATE: 1998-12-28
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 21
; LENGTH: 871
; TYPE: PRT
; ORGANISM: Chlamydia pneumoniae
US-09-886-468-21

Query Match          9.0%; Score 74.5; DB 10; Length 871;
Best Local Similarity 23.9%; Pred. No. 6.4;
Matches 43; Conservative 24; Mismatches 66; Indels 47; Gaps

Qy 22 KPGQLQVNTGSSVLDLSTVTVR---YWFTRD--GGSSTLVYNCDAAMGCG----- 67
Db 307 KTGILAIQSONQEMSFSTNTTTANGACAIYATKCTLDGNTTLTFDONTATACGGGAIYTTETD 366
Qy 68 -NIRAFSGSNPATPTADTYLQL-----SFTGGT---LAAGSGTGEIQNRVHK----- 111
Db 367 FSLKSGTGTVTFSTNTAKTGCALYSKGNSSLTGNLTLLFSGNKATGPSNSSANQEGCGCA 426
Qy 112 -----SDWSNFDETN--DYSYGTNTAFQDW-----TKVTYVYNGRLVNGTSPSGT 154
Db 427 ILAFIDSGVSDKTGLSIANNQEVSLTSNAATVSGGAIYATKCTLTCNGSLTFDGNCTACT 486

RESULT 12
US-09-764-860-377
; Sequence 377, Application US/09764860
; Patent No. US20020094953A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC008
; CURRENT APPLICATION NUMBER: US/09/764,860
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 1198
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 377
; LENGTH: 69
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-764-860-377

Query Match          8.7%; Score 72.5; DB 10; Length 69;
Best Local Similarity 40.9%; Pred. No. 0.4;
Matches 18; Conservative 3; Mismatches 12; Indels 11; Gaps

Qy 30 TGSSSVDLSTVTVTYWFTRDGSGSTLVYNCDAAMGCG---GNTR 70
Db 26 TPSSSSCSNAARYY-----TLKYHCDWFRDCKMWSGTVR 61

RESULT 13
US-09-815-242-10936
; Sequence 10936, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carl, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: Prokaryotes
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078

```

```

: PRIOR FILING DATE: 2000-03-21
: PRIOR APPLICATION NUMBER: 60/206,848
: PRIOR FILING DATE: 2000-05-23
: PRIOR APPLICATION NUMBER: 60/207,727
: PRIOR FILING DATE: 2000-05-26
: PRIOR APPLICATION NUMBER: 60/242,578
: PRIOR FILING DATE: 2000-10-23
: PRIOR APPLICATION NUMBER: 60/253,625
: PRIOR FILING DATE: 2000-11-27
: PRIOR APPLICATION NUMBER: 60/257,931
: PRIOR FILING DATE: 2000-12-22
: PRIOR APPLICATION NUMBER: 60/269,308
: PRIOR FILING DATE: 2001-02-16
: NUMBER OF SEQ ID NOS: 14110
: SOFTWARE: FastSeq for Windows Version 4.0
: SEQ ID NO 10936
: LENGTH: 1333
: TYPE: PRT
: ORGANISM: Enterococcus faecalis
US-09-815-242-10936

Query Match
Best Local Similarity 23.8%; Score 72.5; DB 10; Length 1333;
Matches 34; Conservative 21; Mismatches 65; Indels 23; Gaps 6;

QY 3 GGVKVOYKNNDSAPGD-----NQIKPGLQLVNTGSSVDLSTVTRYWFTRDGGSST 54
      ||:: || || || || || || || || || || || || || || || || || || ||
DB 1021 GGIADIKQPIKVGDIIVLPGNSI---AQIOVTGAQVKEMFEXSVRSIPKDNCTI 1077
      ||:: || || || || || || || || || || || || || || || || || || ||

QY 55 LVYNCOWAMGC-----GNIRASFGSVNPNATPTADTYLQLSFTGGTLAAGSGTG-EI 105
      ||:: || || || || || || || || || || || || || || || || || || ||
DB 1078 LLDDAGQPKLGANGGFLHVSIRIHYDTPKGTRLASD--EGNETGQTIVGSRVLGIEI 1135
      ||:: || || || || || || || || || || || || || || || || || || ||

QY 106 QNRVNSDWSNFDNDYSYGTN 128
      ||:: || || || || || || || || || || || || || || || || || || ||
DB 1136 KNRQTK-LEPLDEKKQYRMATN 1157
      ||:: || || || || || || || || || || || || || || || || || || ||

RESULT 14
US-09-870-521-4
: Sequence 4, Application US/09870521
: Patent No. US20020051989A1
: GENERAL INFORMATION:
: APPLICANT: Miller, Marcia
: APPLICANT: Goto, Ronald
: TITLE OF INVENTION: METHOD FOR BREEDING AND GENOTYPING CHICKENS AND PROBES THEREFOR
: FILE REFERENCE: 1954-310
: CURRENT APPLICATION NUMBER: US/09/870,521
: CURRENT FILING DATE: 2001-06-01
: PRIOR APPLICATION NUMBER: US 60/208471
: PRIOR FILING DATE: 2000-06-02
: NUMBER OF SEQ ID NOS: 11
: SOFTWARE: PatentIn version 3.1
: SEQ ID NO 4
: LENGTH: 334
: TYPE: PRT
: ORGANISM: Gallus sp.
US-09-870-521-4

Query Match
Best Local Similarity 22.6%; Score 70.5; DB 10; Length 334;
Matches 42; Conservative 20; Mismatches 71; Indels 53; Gaps 11;

QY 1 VSGGVKVOYKNNDSAPGD-----NQIKPGLQLVNTGSSVDLSTVTRYWFTRDGGSST 48
      ||:: || || || || || || || || || || || || || || || || || || ||
DB 28 VQGLFVHYNSTARRYPVTEWIAKAQEQYDTGTQKIGGNGRQIDRELNCIPQRY-NKQ 86
      ||:: || || || || || || || || || || || || || || || || || || ||

QY 49 DGSSTL--VYNCOWAMGCCNIRASFGSVNPNATPTADTYLQLSFTG-----GTLLA 98
      ||:: || || || || || || || || || || || || || || || || || || ||
DB 87 TGSHTVQWMYGCD--ILEGGPIRG-----YYQMayDGRDPTAFDKGTMT- 129
      ||:: || || || || || || || || || || || || || || || || || || ||

QY 99 GGSSTGEIQNRV-NKSDWSNFDNDYSYGTNTAFQDKTKVTY-VYNCRL-----VW 147
      ||:: || || || || || || || || || || || || || || || || || || ||
```

```

DB 130 --FTAAVPEAVPTKRKWESEPERKWKNYLEETCTVEWLLRRYVEYVGAELGRRERPEVRVW 187
QY 148 GTEPSG 153
      ||:: || || || || || || || || || || || || || || || || || || ||
DB 188 GREADG 193
      ||:: || || || || || || || || || || || || || || || || || || ||

RESULT 15
US-09-815-242-11359
: Sequence 11359, Application US/09815242
: Patent No. US20020061569A1
: GENERAL INFORMATION:
: APPLICANT: Haselbeck, Robert
: APPLICANT: Ohlsen, Kari L.
: APPLICANT: Zyskind, Judith W.
: APPLICANT: Wall, Daniel
: APPLICANT: Trawick, John D.
: APPLICANT: Carr, Grant J.
: APPLICANT: Yamamoto, Robert T.
: APPLICANT: Xu, H. Howard
: TITLE OF INVENTION: Identification of Essential Genes in
: TITLE OF INVENTION: Prokaryotes
: FILE REFERENCE: ELITRA.011A
: CURRENT APPLICATION NUMBER: US/09/815,242
: CURRENT FILING DATE: 2001-03-21
: PRIOR APPLICATION NUMBER: 60/191,078
: PRIOR FILING DATE: 2000-03-21
: PRIOR APPLICATION NUMBER: 60/206,848
: PRIOR FILING DATE: 2000-05-23
: PRIOR APPLICATION NUMBER: 60/207,727
: PRIOR FILING DATE: 2000-05-26
: PRIOR APPLICATION NUMBER: 60/242,578
: PRIOR FILING DATE: 2000-10-23
: PRIOR APPLICATION NUMBER: 60/253,625
: PRIOR FILING DATE: 2000-11-27
: PRIOR APPLICATION NUMBER: 60/257,931
: PRIOR FILING DATE: 2000-12-22
: PRIOR APPLICATION NUMBER: 60/269,308
: PRIOR FILING DATE: 2001-02-16
: NUMBER OF SEQ ID NOS: 14110
: SOFTWARE: FastSeq for Windows Version 4.0
: SEQ ID NO 11359
: LENGTH: 300
: TYPE: PRT
: ORGANISM: Helicobacter pylori
US-09-815-242-11359

Query Match
Best Local Similarity 22.4%; Score 70; DB 10; Length 300;
Matches 30; Conservative 17; Mismatches 45; Indels 42; Gaps 5;

QY 21 IKPGLQLVNTGSSVDLSTVTRYWFTRDGGSSTLVYNCOWAMGCCNIRASFGSVNPNAT 80
      ||:: || || || || || || || || || || || || || || || || || || ||
DB 90 VDPG---VGTKRKSVLTKNGQYFVSPDNGTLTLVAQ-----TLGIDSVREIDEKANRLK 142
      ||:: || || || || || || || || || || || || || || || || || || ||

QY 81 PTADTYL-----QLSFTGGTLAAGGST-----GEIQNRVVK 111
      ||:: || || || || || || || || || || || || || || || || || || ||
DB 143 GSEKSYTFHGRDYYAYTGARLASGAITFEQVGPPLPKVVEIPIYOKAKATKGEVKGNIPI 202
      ||:: || || || || || || || || || || || || || || || || || || ||

QY 112 SD-----WSNFEDE 119
      ||:: || || || || || || || || || || || || || || || || || || ||
DB 203 LDIQYGNVWSNISD 216
      ||:: || || || || || || || || || || || || || || || || || || ||

Search completed: November 13, 2002, 12:08:36
Job time : 5.98707 secs
```

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run On: November 13, 2002, 11:50:41 : Search time 4.71624 Seconds
(without alignments)
2058.756 Million cell updates/sec

Title: US-09-917-378-5

Perfect score: 543

Sequence: 1 GVGCRATVYVNSDWSGFTA.....GFNGSYSGTNTAPTTLCTAS 101

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : PIR_73.*

1: pir1.*

2: pir2.*

3: pir3.*

4: pir4.*

Pred. No. Is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	292.5	53.9	579	2 T35240	probable secreted
2	275	50.6	596	2 A55976	cellulose 1,4-beta
3	270	49.7	441	2 T12011	cellulase (EC 3.2.
4	266.5	49.1	747	2 B47093	cellulose 1,4-beta
5	265	48.8	984	2 T44496	cellulose 1,4-beta
6	257	47.3	890	2 T35237	probable secreted
7	254	46.8	495	2 T17478	hypothetical prote
8	244.5	45.0	484	1 A24994	cellulose 1,4-beta
9	236.5	43.6	872	2 S49541	cellulase - Cellul
10	234	43.1	1045	2 A39199	endoglucanase B (E
11	230	42.4	748	2 S19652	cellodextrinase C
12	228.5	42.1	973	2 T35238	probable secreted
13	223.5	41.2	1090	2 S59077	cellulose 1,4-beta
14	213	39.2	609	2 T42073	probable chitinase
15	212.5	39.1	449	2 A24993	cellulase (EC 3.2.
16	212.5	39.1	515	2 S20493	cellulase (EC 3.2.
17	205	37.8	610	2 JH0573	endoglucanase - Cl
18	198	36.5	570	2 S56132	chitinase (EC 3.2.
19	197	36.3	400	2 T31066	cellulase (EC 3.2.
20	197	36.3	426	2 A42360	probable secreted
21	192	35.4	962	2 S03818	cellulase (EC 3.2.
22	191.5	35.3	382	2 JH0571	carboxymethylcellu
23	184.5	34.0	571	2 S13392	cellulase (EC 3.2.
24	184.5	34.0	583	2 S36781	alpha-N-arabinofur
25	184.5	34.0	592	1 S13391	esterase D - Pseud
26	184	33.9	611	1 S06047	endo-1,4-beta-xyla
27	180	33.1	511	1 S10527	endo-1,4-beta-xyla
28	171	31.5	466	2 C42360	endoglucanase B pr
29	155	28.5	142	2 A70757	cellulase (EC 3.2.
					hypothetical prote

30 138.5 25.5 644 1 I40712
31 137 25.2 683 2 A82704
32 132.5 24.4 592 2 E82759
33 122.5 22.6 547 2 J00356
34 121 22.3 1070 2 S75712
35 116.5 21.5 765 2 T35719
36 104 19.2 995 2 S50358
37 97 17.9 1778 2 T50074
38 94 17.3 335 2 T50601
39 93 17.1 1026 2 E86185
40 93 17.1 1026 2 A48995
41 93 17.1 1073 2 C87374
42 89.5 16.5 237 2 S12610
43 88 16.2 335 2 T50600
44 88 16.2 717 2 T06041
45 87.5 16.1 1034 2 JC2143

ALIGNMENTS

RESULT 1

T35240

probable secreted cellulase - Streptomyces coelicolor

C:Species: Streptomyces coelicolor

C:Date: 05-Nov-1999 #sequence_revision 05-Nov-1999 #text_change 05-Nov-1999

C:Accession: T35240

R:Seeger, K.J.; Harris, D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A.

submitted to the EMBL Data Library, September 1998

A:Reference number: 221572

A:Accession: T35240

A:Status: preliminary; translated from GB/EMBL/DDBJ

A:Molecule type: DNA

A:Residues: 1-579 <SEE>

A:Cross-references: EMBL:AL031515; PIDN:CAA20645.1; GSPDB:GN00070; SCOEDB:SC5C7.33

A:Experimental source: strain A3(2)

C:Genetics:

A:Gene: SCOEDB:SC5C7.33

Query Match 53.9%; Score 292.5; DB 2; Length 579;

Best Local Similarity 53.5%; Pred. No. 3e-18;

Matches 54; Conservative 14; Mismatches 26; Indels 7; Gaps 2;

Qy 4 CRATVYVNSDWSGFTAATVTNTGSRATSGTVAWSFGGNOTVTNYNTALTQSGASVT 63

Db 37 CTVDYQVNDWGSFGTAATVTNNGA-ATSSMSLGTAGSQKVTNSWNAKVTOSCAAVT 95

Qy 64 ATNLSYNNVIQPGOSTTFGFGNGSYSGTNTAPT-----LTC 98

Db 96 AANESYNTLTSGSASFGFGTGYSGSNAIPATFTLNGVTC 136

RESULT 2

A55976

cellulose 1,4-beta-cellobiosidase (EC 3.2.1.91) - Thermomonospora fusca

C:Species: Thermomonospora fusca

C:Date: 05-Jan-1996 #sequence_revision 05-Jan-1996 #text_change 22-Oct-1999

R:Zhang, S.; Lao, G.; Wilson, D.B.

Biochemistry 34, 3386-3395, 1995

A:Title: Characterization of a Thermomonospora fusca exocellulase.

A:Reference number: A55976; MUID:95186496; PMID:7880834

A:Accession: A55976

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-596 <2HA>

A:Cross-references: EMBL:U18978; NID:g664822; PIDN:AAA62211.1; PID:g664823

C:Superfamily: bacterial cellulose-binding domain homology

C:Keywords: glycosidase; hydrolase; polysaccharide degradation

F:40-140/Domain: bacterial cellulose-binding domain homology <BCB>

F:41-139/Disulfide bonds: #status predicted

Query Match 50.6%; Score 275; DB 2; Length 596;

Best Local Similarity 52.4%; Pred. No. 1.le-16;
Matches 55; Conservative 10; Mismatches 32; Indels 8; Gaps 3;

Qy- 3 GCRTATVYVNSDMGSGFTATVTNTGSRATSGMTVAWSFGGNOTVTYNTALTQSGASV 62
Db 40 GCSVDYTVNS-WGTFGTANTITNLGS-AINGMTLEWDFPGNQVNTLWNGTYTQSGOHV 97

Qy 63 TATNLSYNNVIOQSGSTTFGNGSYSGTNTAPT-----LTCTAS 101
Db 98 SVSNAPYNASIPANGTVEFGNGSYSGSNDIPSSFKLNGVTCDGS 142

RESULT 3

T12011
cellulase (EC 3.2.1.4) precursor - Thermomonospora fusca
N:Alternate names: endo-1,4-beta-glucanase
C:Species: Thermomonospora fusca
C>Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 16-Jul-1999
C:Accession: T12011
R:Zhang, S.; Wilson, D.B.
J. Biotechnol. 57, 101-113, 1997
A:Title: Surface residue mutations which change the substrate specificity of Thermomonospora fusca
A:Reference number: Z17376; MUID:97475714; PMID:9335169
A:Accession: T12011
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-441 <ZHA>
A:Cross-references: EMBL:M73321; NID:q436311; PID:q436312
C:Keywords: glycosidase; hydrolase; polysaccharide degradation

Query Match 49.7%; Score 270; DB 2; Length 441;
Best Local Similarity 50.0%; Pred. No. 2.le-16;
Matches 49; Conservative 16; Mismatches 31; Indels 2; Gaps 1;

Qy 4 CRATYVYVNSDMGSGFTATVTNTGSRATSGMTVAWSFGGNOTVTYNTALTQSGASV 63
Db 346 CTATYTIANENWDFQATVTV-TANQITGWTVTWITTCQTITNANADVSTSGSSVT 403

Qy 64 ATNLNSYNNVIOQSGSTTFGNGSYSGTNTAPTTLCTAS 101
Db 404 ARNVGHGNTLSQGAATSTFGFVSGKSGNSVPTLTCAAS 441

RESULT 4

B47093
cellulase (EC 3.2.1.4) CenD - Cellulomonas fimi
N:Alternate names: endo-1,4-beta-glucanase
C:Species: Cellulomonas fimi
C>Date: 21-Sep-1993 #sequence_revision 18-Nov-1994 #text_change 06-Dec-1996
C:Accession: B47093
R:Meinke, A.; Glikes, N.R.; Kilburn, D.G.; Miller Jr., R.C.; Warren, R.A.
J. Bacteriol. 175, 1910-1918, 1993
A:Title: Cellulose-binding polypeptides from Cellulomonas fimi: endoglucanase D (CenD),
A:Reference number: A47093; MUID:93209933; PMID:8458833
A:Accession: B47093
A:Status: preliminary
A:Molecule type: DNA; protein
A:Residues: 1-747 <MEI>
A:Experimental source: ATCC 484
A:Note: sequence inconsistent with the nucleotide translation
A:Note: sequence extracted from NCBI backbone (NCBIN:128120, NCBIPI:128122)
C:Function:
A:Description: hydrolysis of 1,4-beta-D-glucosidic linkages in beta-D-glucans such as cel-
lulose
A:Pathway: cellulose degradation
C:Superfamily: bacterial cellulose-binding domain homology
C:Keywords: glycosidase; hydrolase; polysaccharide degradation
F:644-745/Domain: bacterial cellulose-binding domain homology <BCB>
F:645-744/Disulfide bonds: #status predicted

Query Match 49.1%; Score 266.5; DB 2; Length 747;
Best Local Similarity 50.0%; Pred. No. 7.6e-16;
Matches 51; Conservative 15; Mismatches 29; Indels 7; Gaps 2;

Qy 4 CRATYVYVNSDMGSGFTATVTNTGSRATSGMTVAWSFGGNOTVTYNTALTQSGASV 63
Db 645 CAVTYTANG-WSGGFTAAVTLTNTGTTALSCWTLGFPAPPSGOTLTGCHSAHWAQSGSSVT 703

Qy 64 ATNLNSYNNVIOQSGSTTFGNGSYSGTNTAPT-----LTCT 99
Db 704 ATNEAMNVLAPGASVEIGFSGTITGNTAPATFTVGGATCT 745

RESULT 5

T44496
cellulose 1,4-beta-cellobiosidase (EC 3.2.1.91) precursor [Imported] - Thermomonospora
N:Alternate names: beta-1,4-exocellulase E6 precursor
C:Species: Thermomonospora fusca
C>Date: 21-Jan-2000 #sequence_revision 21-Jan-2000 #text_change 21-Jan-2000
C:Accession: T44496
R:Irwin, D.I.; Zhang, S.; Wilson, D.B.
submitted to the EMBL Data Library, April 1999
A:Description: Characterization of a Thermomonospora fusca family 48 exocellulase E6.
A:Reference number: 222783

A:Accession: T44496
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-984 <IRW>
A:Cross-references: EMBL:AF144563; PIDN:AAD39947.1
A:Experimental source: strain YX
C:Genetics:
A:Gene: celf

C:Keywords: glycosidase; hydrolase
F:1-34/Domain: signal sequence #status predicted <SIG>
F:35-984/Product: cellulose 1,4-beta-cellobiosidase #status predicted <MAT>

Query Match 48.8%; Score 265; DB 2; Length 984;
Best Local Similarity 48.1%; Pred. No. 1.4e-15;
Matches 51; Conservative 13; Mismatches 36; Indels 6; Gaps 1;

Qy 2 VGCRTATVYVNSDMGSGFTATVTNTGSRATSGMTVAWSFGGNOTVTYNTALTQSGAS 61
Db 36 VACSVDYDDSDNDGSGFVAEVKVTNEGSDPIQNNQVCHTFPGNOOITNGWNGVFSOSCAN 95

Qy 62 VTATNLNSYNNVIOQSGSTTFGNGSYSGTNTAPT-----LTCTAS 101
Db 96 VTVRYPDWNPNIAPGATISFGOGTSGSNDAPTSTFTVNGVTCSGS 141

RESULT 6

T35237
probable secreted cellulase - Streptomyces coelicolor
C:Species: Streptomyces coelicolor
C>Date: 05-Nov-1999 #sequence_revision 05-Nov-1999 #text_change 05-Nov-1999
C:Accession: T35237
R:Seeger, K.J.; Harris, D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A.
submitted to the EMBL Data Library, September 1998
A:Reference number: Z21572
A:Accession: T35237
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-890 <SEE>
A:Cross-references: EMBL:AL031515; PIDN:CAA20642.1; GSPDB:GN00070; SCOPDB:SC5C7.30C
A:Experimental source: strain A3(2)
C:Genetics:
A:Gene: SCOPDB:SC5C7.30C

Query Match 47.3%; Score 257; DB 2; Length 890;
Best Local Similarity 51.6%; Pred. No. 6.2e-15;
Matches 47; Conservative 10; Mismatches 34; Indels 0; Gaps 0;

Qy 4 CRATYVYVNSDMGSGFTATVTNTGSRATSGMTVAWSFGGNOTVTYNTALTQSGASV 63
Db 786 CEVTYTVTNQWPGCFQADVRLTNTGTSANWGSWLDHFFPGQGVTRMNAEHTOAGTSVT 845

Qy 64 ATNLNSYNNVIOQSGSTTFGNGSYSGTNTAP 94
Db 64 ATNLNSYNNVIOQSGSTTFGNGSYSGTNTAP 94

A;CROSS-REFERENCES: EMBL:AL031313; PIDN:CAA20643.1; GSPDH:GN0007

R; Salt

R; Saïto, A.; Fujii, T.; Yoneyama, M.; Ohno, T.; Watanabe, T.; Miyashita, T.

Search completed: November 13, 2002, 11:57:36
Job time : 5.71624 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 13, 2002, 11:47:41 ; Search time 2.32184 Seconds
(without alignments)
1804.220 Million cell updates/sec

Title: US-09-917-378-5

Perfect score: 543

Sequence: 1 GVGCRATVYVNSDWSGFTA.....GFNGSYSGTNTAPTTLCTAS 101

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 20000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	399	73.5	562	1	GUN1_ACICE
2	270	49.7	441	1	GUN2_THREPU
3	266.5	49.1	747	1	GUN2_CELFI
4	244.5	45.0	484	1	GUX_CELFI
5	240.5	44.3	456	1	GUNA_MICBI
6	236.5	43.6	872	1	GUXA_CELFI
7	234	43.1	1045	1	GUNB_CELFI
8	230	42.4	748	1	GUNC_PSEFL
9	225.5	41.5	880	1	GUN4_THREPU
10	223.5	41.2	1090	1	GUXB_CELFI
11	213	39.2	619	1	CHIT_STRLI
12	212.5	39.1	449	1	GUNA_CELFI
13	212.5	39.1	515	1	GUNB_CLOCL
14	205	37.8	610	1	CHIT_STRPL
15	202	37.2	459	1	GUNA_STRLI
16	192	35.4	962	1	GUNA_PSEFL
17	184.5	34.0	571	1	XYNC_PSEFL
18	184.5	34.0	592	1	XYNB_PSEFL
19	184	33.9	611	1	XYNA_PSEFL
20	180	33.1	511	1	GUNB_PSEFL
21	171	31.5	466	1	GUN5_THREPU
22	155	28.5	142	1	YJ87_MYCTU
23	138.5	25.5	484	1	GUNA_XANCP
24	138.5	25.5	644	1	XYND_CELFI
25	122.5	22.6	547	1	GUN1_BUTFI
26	104	19.2	995	1	Y109_YEAST
27	96	17.7	1229	1	N121_HUMAN
28	93	17.1	1025	1	SLAP_CAUCL
29	91	16.8	335	1	XYNB_STRLI
30	89.5	16.5	237	1	GUN_ASFAC
31	87.5	16.1	1034	1	ICEN_PANAN
32	87.5	16.1	1258	1	ICEN_ERWHE
33	87.5	16.1	1322	1	ICEA_PANAN

ALIGNMENTS

RESULT 1

ID	GUN1_ACICE	STANDARD;	PRT;	562 AA.
AD	P54583:			
DT	01-OCT-1996 (Rel. 34, Created)			
DT	01-OCT-1996 (Rel. 34, Last sequence update)			
DT	01-OCT-1996 (Rel. 34, Last annotation update)			
DE	Endoglucanase E1 precursor (EC 3.2.1.4) (Endo-1,4-beta-glucanase E1)			
DE	(Cellulase E1) (Endocellulase E1).			
OS	Acidothermus cellulolyticus.			
OC	Bacteria; Actinobacteria; Actinobacterla (class); Actinobacteridae;			
OC	Actinomycetales; Frankineae; Acidothermaceae; Acidothermus.			
OX	NCBI_TaxID=28049;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN-ATCC 43068 / 11B;			
RA	Laymon R.A., Himmel M.E., Thomas S.R.;			
RL	Submitted (AUG-1995) to the EMBL/GenBank/DBJ databases.			
RN	[2]			
RP	X-RAY CRYSTALLOGRAPHY (2.4 ANGSTROMS) OF 42-398.			
RX	MEDLINE=96346058; PubMed=8718854;			
RA	Saxon J., Adney W.S., Himmel M.E., Thomas S.R., Kaplan P.A.;			
RT	"Crystal structure of thermostable family 5 endocellulase E1 from Acidothermus cellulolyticus in complex with cellobiose.;"			
RL	Biochemistry 35:10648-10660(1996).			
CC	-!- FUNCTION: THERMOSTABLE ENZYME WITH AN OPTIMAL TEMPERATURE OF 81			
CC	DEGREES CELSIUS. HAS A VERY HIGH SPECIFIC ACTIVITY ON			
CC	CARBOXYMETHYLCELLULOSE.			
CC	-!- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic			
CC	linkages in cellulose.			
CC	-!- SIMILARITY: BELONGS TO CELLULASE FAMILY A (FAMILY 5 OF GLYCOSYL			
CC	HYDROLASES).			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL Outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/			
CC	or send an email to license@isb-sib.ch).			
CC	-----			
EMBL	U33212; AAA75477.1; -			
PDB	1ECE; 14-OCT-96.			
DR	InterPro: IPR001919; Bac_celose-bind.			
DR	InterPro: IPR001547; GH_5.			
DR	Pfam: PF00150; cellulase; 1.			
DR	Pfam: PF00553; CBM_2; 1.			
DR	PROSITE: PS00659; GLYCOSYL_HYDROL_F5; 1.			
KW	Cellulose degradation; Hydrolase; Glycosidase; Signal; 3D-structure.			
FT	SIGNAL 1 41			
FT	CHAIN 42 562			
FT	DOMAIN 42 400			
FT	DOMAIN 401 461			
FT	DOMAIN 462 562			
FT	ACT_SITE 203 203			
FT	ACT_SITE 323 323			

P54937 clostridium
P52948 homo sapien
O00130 lctalulrid h
O83831 treponema p
P29957 alteromonas
O03155 escherichia
O52530 caulobacter
P22698 dictyostell
O33479 pseudomonas
O04893 saccharomyc
O30611 pseudomonas
P23314 xanthomonas

34 84.5 15.6 517 1 GUNA_CLOLO
35 83.5 15.4 937 1 NU98_HUMAN
36 83 15.3 670 1 VG50_HSV11
37 81.5 15.0 494 1 Y859_TREPA
38 81.5 15.0 669 1 AMY_ALTHA
39 81 14.9 1286 1 AIDA_ECOLI
40 80.5 14.8 273 1 FLJN_CAUCR
41 80.5 14.8 532 1 SPG7_DICDI
42 80.5 14.8 1196 1 ICEV_PSESX
43 79.5 14.6 1140 1 YM96_YEAST
44 79.5 14.6 1148 1 ICEK_PSESX
45 78.5 14.5 580 1 PROA_XANCP

```
FT DISULFID 75 161
FT DISULFID 209 212
SQ SEQUENCE 562 AA: 60747 MW: 84E6256406A35041 CRC64:

Query Match 73.5% Score 399: DB 1: Length 562;
Best Local Similarity 69.3%; Pred. No. 2.3e-28;
Matches 70; Conservative 13; Mismatches 18; Indels 0; Gaps 0;

Qy 1 GVCGRATVYVNSDMGSGFTATVTNTGSRATSCWTVAWSFGGNOTVTNYNTALTQSGA 60
   I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I
Db 462 GARTASYOVNSDNGNGFTVTAVTNSGVSATKTWTVSWTEGGNGQTITNSNAAVTQNGQ 521
   I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I
Qy 61 SVTATNLISYNNVIQPGOSTTGFNGSGTNTAPLTCTAS 101
   I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I
Db 522 SVTARNNSYNNVIQPGONTTGFQASVTGTSNAAPTACAAAS 562
   I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I

RESULT 2
GUN2-THEFU STANDARD: PRT: 441 AA.
AC P26222;
DT 01-MAY-1992 (Rel. 22, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Endoglucanase E-2 precursor (EC 3.2.1.4) (Endo-1,4-beta-glucanase E-2)
DE (Cellulase E-2) (Cellulase E2).
GN CELB.
OS Thermomonospora fusca.
OC Bacteria; Actinobacteria; Actinobacteria (class); Actinobacteridae;
OC Actinomycetales; Streptosporangineae; Nocardiopsaceae; Thermobifida.
OX NCBI_TaxID=2021;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=YX;
RX MEDLINE=91258320; PubMed=1904434;
RA Lao G., Ghangas G.S., Jung E.D., Wilson D.B.;
RT "DNA sequences of three beta-1,4-endoglucanase genes from
RL Thermomonospora fusca.";
RN J. Bacteriol. 173:3397-3407(1991).
RN [2]
RP REVISIONS, SEQUENCE FROM N.A.
RC STRAIN=YX;
RA Jung E.D., Lao G., Irwin D., Barr B., Benjamin A., Wilson D.B.;
RL Submitted (MAY-1993) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE OF 32-47.
RA Wilson D.B.;
RT "Cellulases of Thermomonospora fusca.";
RL Meth. Enzymol. 160:314-323(1988).
RN [4]
RP X-RAY CRYSTALLOGRAPHY (1.8 ANGSTROMS) OF 32-317.
RX MEDLINE=94002001; PubMed=8399160;
RA Spezio M., Wilson D.B., Karplus P.A.;
RT "Crystal structure of the catalytic domain of a thermophilic
RL endocellulase.";
RL Biochemistry 32:9906-9916(1993).
CC -1- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic
CC linkages in cellulose.
CC -1- PATHWAY: Cellulose degradation.
CC -1- SIMILARITY: CONTAINS 1 BACTERIAL-TYPE CELLULOSE-BINDING DOMAIN
CC (CBD).
CC -1- SIMILARITY: BELONGS TO CELLULOSE FAMILY B (FAMILY 6 OF GLYCOSYL
CC HYDROLASES).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@lsb-sib.ch).
CC -----
DR EMBL; M73321; AAC06388.1; -
```

```
DR PIR: A42360; A42360.
DR PDB: 1TML; 31-JAN-94.
DR InterPro: IPR001919; Bac_cellose-bind.
DR InterPro: IPR001524; GH_6.
DR Pfam: PF00553; CBM_2; 1.
DR PRINTS: PR00733; GLHYDROLASE6.
DR PRODOM: PD003733; GH_6; 1.
DR PROSITE: PS00561; CBD_BACTERIAL; 1.
DR PROSITE: PS00655; GLYCOSYL_HYDROL_F6_1; 1.
DR PROSITE: PS00656; GLYCOSYL_HYDROL_F6_2; 1.
KW Cellulose degradation; Hydrolase; Glycosidase; Signal; 3D-structure.
FT SIGNAL 1 31
FT CHAIN 32 441 ENDOGLUCANASE E-2.
FT DOMAIN 32 320 CATALYTIC.
FT DOMAIN 321 340 LINKER.
FT DOMAIN 341 441 CELLULOSE-BINDING.
FT ACT_SITE 110 110
FT ACT_SITE 148 148
FT ACT_SITE 296 296
FT DISULFID 111 156
FT DISULFID 263 298
FT DISULFID 346 438
FT STRAND 36 36
FT TURN 40 41
FT HELIX 43 50
FT TURN 52 53
FT TURN 55 56
FT HELIX 57 63
FT TURN 64 66
FT STRAND 69 69
FT STRAND 71 73
FT HELIX 78 95
FT TURN 96 96
FT STRAND 98 98
FT STRAND 100 103
FT TURN 109 112
FT HELIX 122 134
FT TURN 135 138
FT STRAND 142 145
FT TURN 147 148
FT HELIX 149 153
FT TURN 154 155
FT HELIX 158 178
FT TURN 180 181
FT STRAND 183 187
FT HELIX 196 205
FT TURN 206 207
FT STRAND 214 217
FT TURN 219 220
FT HELIX 225 239
FT TURN 240 240
FT TURN 242 243
FT STRAND 245 249
FT TURN 254 255
FT TURN 259 260
FT TURN 266 267
FT STRAND 275 275
FT TURN 281 282
FT STRAND 283 288
FT TURN 292 293
FT STRAND 294 294
FT TURN 302 303
FT STRAND 305 305
FT HELIX 307 315
FT TURN 316 316
SQ SEQUENCE 441 AA: 45843 MW: 87218E4537092AE5 CRC64:

Query Match 49.7% Score 270; DB 1: Length 441;
Best Local Similarity 50.0%; Pred. No. 4.8e-17;
Matches 49; Conservative 16; Mismatches 31; Indels 2; Gaps 1;
```

QY 4 CRATVYVNSDMGSGFTATVTVTNTGSRATSGTWAVSEGGNQTVTNYWNTALTOSGASVT 63
 | | | | : : | | | | | : : | | | | | : : | | | | | : : | | | | |
 Db 346 CTATVTIANENWDGFOATVTV--TANQNITGWTVTWTFDDGQTIITNANNADVTSQSSVT 403
 | | : : | | | | | | : : | | | | | : : | | | | | : : | | | | |
 QY 64 ATNLSYNNVIOGOSTTFCFNGCSYSGTWTAPLTCTAS 101
 | : : | : | | | | | | : : | | | | | : : | | | | | : : | | | | |
 Db 404 ARNVGHNGTLSOGASTFEGVCGSKGNSNSVPTLTCAAS 441
 | : : | : | | | | | | : : | | | | | : : | | | | | : : | | | | |

RESULT 3
 GUND_CELFI
 ID GUND_CELFI STANDARD: PRT: 747 AA.
 AC P50400:
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Endoglucanase D precursor (EC 3.2.1.4) (Endo-1,4-beta-glucanase)
 DE (Cellulase).
 GN CEND.
 OS Cellulomonas fimi.
 OC Bacteria; Actinobacteria; Actinobacteria (class); Actinobacteridae;
 OC Actinomycetales; Micrococcales; Cellulomonadaceae; Cellulomonas.
 OX NCBI_TaxID=1708;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=93209933; PubMed=8458833;
 RA Melnke A., Gilkes N.R., Kilburn D.G., Miller R.C. Jr., Warren R.A.;
 RT "Cellulose-binding polypeptides from Cellulomonas fimi: endoglucanase
 D (Cend), a family A beta-1,4-glucanase.";
 RL J. Bacteriol. 175:1910-1918(1993).
 CC -|- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic
 linkages in cellulose.
 CC -|- PATHWAY: Cellulose degradation.
 CC -|- SIMILARITY: CONTAINS 2 FIBRONECTIN TYPE III-LIKE DOMAINS.
 CC -|- SIMILARITY: CONTAINS 1 BACTERIAL-TYPE CELLULOSE-BINDING DOMAIN
 (CBD).
 CC -|- SIMILARITY: BELONGS TO CELLULOSE FAMILY A (FAMILY 5 OF GLYCOSYL
 HYDROLASES).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: L02544; AAA23089.1; -;
 DR HSP; P07986; IEXG.
 DR InterPro: IPR001919; Bac_celose-bind.
 DR InterPro: IPR003961; FN_III.
 DR InterPro: IPR003962; FNIII_repeat.
 DR InterPro: IPR001547; GH_5.
 DR Pfam: PF00041; fn3; 2
 DR Pfam: PF00150; cellulase; 1.
 DR Pfam: PF00553; CBM2; 1.
 DR PRINTS: PR00014; FNTYPEIII.
 DR SMART: SM00060; FN3; 2.
 DR PROSITE: PS00659; GLYCOSYL_HYDROL_F5; 1.
 DR Cellulose degradation; Hydrolase; Glycosidase; Repeat; Signal.
 FT SIGNAL 1 39 POTENTIAL.
 FT CHAIN 40 747 ENDOGLUCANASE D.
 FT DOMAIN 448 542 FIBRONECTIN TYPE-III 1.
 FT DOMAIN 546 639 FIBRONECTIN TYPE-III 2.
 FT ACT_SITE 208 208 PROTON DONOR (BY SIMILARITY).
 FT ACT_SITE 349 349 NUCLEOPHILE (BY SIMILARITY).
 SQ SEQUENCE 747 AA; 79936 MW; BD15473C9DBB42BD CRC64;

Query Match 49.1%; Score 266.5; DB 1; Length 747;
 Best Local Similarity 50.0%; Pred. No. 1.7e-16;
 Matches 51; Conservative 15; Mismatches 29; Indels 7; Gaps 2;

QY 4 CRATVYVNSDMGSGFTATVTVTNTGSRATSGTWAVSEGGNQTVTNYWNTALTOSGASVT 63
 | | | | : : | | | | | : : | | | | | : : | | | | | : : | | | | |
 Db 346 CTATVTIANENWDGFOATVTV--TANQNITGWTVTWTFDDGQTIITNANNADVTSQSSVT 403
 | | : : | | | | | | : : | | | | | : : | | | | | : : | | | | |
 QY 64 ATNLSYNNVIOGOSTTFCFNGCSYSGTWTAPLTCTAS 101
 | : : | : | | | | | | : : | | | | | : : | | | | | : : | | | | |
 Db 404 ARNVGHNGTLSOGASTFEGVCGSKGNSNSVPTLTCAAS 441
 | : : | : | | | | | | : : | | | | | : : | | | | | : : | | | | |

RESULT 4
 GUX_CELFI
 ID GUX_CELFI STANDARD: PRT: 484 AA.
 AC P07986;
 DT 01-AUG-1988 (Rel. 08, Created)
 DT 01-AUG-1988 (Rel. 08, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Exoglucanase/xylanase precursor [includes: Exoglucanase (EC 3.2.1.91)
 DE (Exocellulohydrolase) (1,4-beta-cellobiohydrolase) (beta-1,4-
 DE glycanase CEX); Endo-1,4-beta-xylanase B (EC 3.2.1.8) (Xylanase B)].
 GN CEX OR XYNB.
 OS Cellulomonas fimi.
 OC Bacteria; Actinobacteria; Actinobacteria (class); Actinobacteridae;
 OC Actinomycetales; Micrococcales; Cellulomonadaceae; Cellulomonas.
 OX NCBI_TaxID=1708;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=87055250; PubMed=3096818;
 RA O'Neill G., Goh S.H., Warren R.A.J., Kilburn D.G., Miller R.C. Jr.;
 RT "Structure of the gene encoding the exoglucanase of Cellulomonas
 fimi.";
 RL Gene 44:325-330(1986).
 RN [2]
 RP ACTIVE SITE GLU-274.
 RX MEDLINE=91340691; PubMed=1678739;
 RA Tull D., Withers S.G., Gilkes N.R., Kilburn D.G., Warren R.A.J.,
 RA Aebersold R.;
 RT "Glutamic acid 274 is the nucleophile in the active site of a
 'retaining' exoglucanase from Cellulomonas fimi.";
 RL J. Biol. Chem. 266:15621-15625(1991).
 RN [3]
 RP DISULFIDE BONDS.
 RX MEDLINE=92104156; PubMed=1761039;
 RA Gilkes N.R., Claeysens M., Aebersold R., Henrissat B., Melnke A.,
 RA Morrison H.D., Kilburn D.G., Warren R.A.J., Miller R.C. Jr.;
 RT "Structural and functional relationships in two families of beta-1,4-
 RT glycanases.";
 RL Eur. J. Biochem. 202:367-377(1991).
 RN [4]
 RP X-RAY CRYSTALLOGRAPHY (1.8 ANGSTROMS).
 RX MEDLINE=95001978; PubMed=7918478;
 RA White A., Withers S.G., Gilkes N.R., Rose D.R.;
 RT "Crystal structure of the catalytic domain of the beta-1,4-glycanase
 cex from Cellulomonas fimi.";
 RL Biochemistry 33:12546-12552(1994).
 RN [5]
 RP X-RAY CRYSTALLOGRAPHY (1.8 ANGSTROMS).
 RX MEDLINE=96163434; PubMed=8564541;
 RA White A., Tull D., Johns K., Withers S.G., Rose D.R.;
 RT "Crystallographic observation of a covalent catalytic intermediate in
 RT a beta-glycosidase.";
 RL Nat. Struct. Biol. 3:149-154(1996).
 RN [6]
 RP X-RAY CRYSTALLOGRAPHY (1.9 ANGSTROMS) OF 42-353.
 RX MEDLINE=98206890; PubMed=9537990;
 RA Notenboom V., Birsan C., Warren R.A.J., Withers S.G., Rose D.R.;
 RT "Exploring the cellulose/xylan specificity of the beta-1,4-glycanase
 cex from Cellulomonas fimi through crystallography and mutation.";
 RL Biochemistry 37:4751-4758(1998).
 RN [7]
 RP X-RAY CRYSTALLOGRAPHY (1.8 ANGSTROMS) OF 42-353.
 RX STRAIN=ATCC 484;
 RA MEDLINE=98400502; PubMed=9731776;
 RA Notenboom V., Birsan C., Niltz M., Rose D.R., Warren R.A.,
 RA Withers S.G.;

RT "Insights into transition state stabilization of the beta-1,4-glycosidase Cex by covalent intermediate accumulation in active site mutants."; Nat. Struct. Biol. 5:812-818(1998).

RN [9]

RP STRUCTURE BY NMR OF 377-484.

RX MEDLINE-95284032; PubMed-7766609;

RA Xu G.-Y., Ong E., Gilkes N.R., Kilburn D.G., Muhandiram D.R., Harris-Brandts M., Carver J.P., Kay L.E., Harvey T.S.; "Solution structure of a cellulose-binding domain from Cellulomonas fimi by nuclear magnetic resonance spectroscopy."; Biochemistry 34:6993-7009(1995).

RN [9]

RP MUTAGENESIS OF GLU-168.

RX MEDLINE-94250681; PubMed-7910761;

RA MacLeod A.M., Lindhorst T., Withers S.G., Warren R.A.J.; "The acid/base catalyst in the exoglucanase/xylanase from Cellulomonas fimi is glutamic acid 127: evidence from detailed kinetic studies of mutants."; Biochemistry 33:6371-6376(1994).

RL -1- FUNCTION: HYDROLYSES BOTH CELLULOSE AND XYLAN. HAS ALSO WEAK ENDOGLUCANASE ACTIVITY.

CC -1- FUNCTION: THE BIOLOGICAL CONVERSION OF CELLULOSE TO GLUCOSE GENERALLY REQUIRES THREE TYPES OF HYDROLYTIC ENZYMES:

CC (1) ENDOGLUCANASES WHICH CUT INTERNAL BETA-1,4-GLUCOSIDIC BONDS;

CC (2) EXOCELLULOBIODHOLASES THAT CUT THE DISSACCHARIDE CELLULOSE FROM THE NONREDUCING END OF THE CELLULOSE POLYMER CHAIN;

CC (3) BETA-1,4-GLUCOSIDASES WHICH HYDROLYZE THE CELLULOSE AND OTHER SHORT CELLO-OLIGOSACCHARIDES TO GLUCOSE.

CC -1- CATALYTIC ACTIVITY: Hydrolysis of 1,4-beta-D-glucosidic linkages in cellulose and cellotetraose, releasing cellobiose from the non-reducing ends of the chains.

CC -1- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-xylosidic linkages in xylans.

CC -1- MISCELLANEOUS: THE LINKER REGION (ALSO TERMED "HINGE") MAY BE A POTENTIAL SITE FOR PROTEOLYSIS

CC -1- SIMILARITY: BELONGS TO CELLULOSE FAMILY F (FAMILY 10 OF GLYCOSYL HYDROLASES).

CC -1- SIMILARITY: CONTAINS 1 BACTERIAL-TYPE CELLULOSE-BINDING DOMAIN (CBD).

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC -----

DR ENBL: M15824; AA56791.1; -

DR PIR: A24994; A24994.

DR PDB: 2EXO; 08-MAR-95.

DR PDB: 1EXG; 03-JUN-95.

DR PDB: 1EXH; 03-JUN-95.

DR PDB: 1EXP; 27-JAN-97.

DR PDB: 2XYL; 18-MAR-98.

DR PDB: 2HIS; 14-OCT-98.

DR InterPro: IPR001919; Bac_cellose-bind.

DR InterPro: IPR001000; Glyco_hydro_10.

DR InterPro: IPR001230; Prenyl_site.

DR Pfam: PF00331; Glyco_hydro_10; 1.

DR Pfam: PF00553; CBM_2; 1.

DR PRINTS: PR00134; GLYDRLASE10.

DR PROSITE: PS00561; CBD_BACTERIAL; 1.

DR PROSITE: PS00591; GLYCOSYL_HYDROL_F10; 1.

KW Cellulose degradation; Hydrolase; Glycosidase; Repeat; Signal;

KW Xylan degradation; Multifunctional enzyme; 3D-structure.

FT SIGNAL 1 41

FT CHAIN 42 484 EXOGLUCANASE/XYLANASE.

FT DOMAIN 42 356 CATALYTIC.

FT DOMAIN 357 376 LINKER ("HINGE") (PRO-THR BOX).

FT DOMAIN 377 484 CELLULOSE-BINDING (BY SIMILARITY).

FT ACT_SITE 168 168 PROTON DONOR.

FT ACT_SITE 274 274 NUCLEOPHILE.

FT DISULFID 208 240

FT DISULFID 302 308

FT DISULFID 382 481

FT MUTAGEN 168 168 E->A,D,G: REDUCED ACTIVITY.

SQ SEQUENCE 484 AA; 51291 MW; 6EE5486BC0E9B02F CRC64;

Query Match 45.0%; Score 244.5; DB 1; Length 484;

Best Local Similarity 49.5%; Pred. No. 9.6e-15;

Matches 46; Conservative 14; Mismatches 32; Indels 1; Gaps 1;

QY 3 GCRTATVYVNSDWGSGTATVTNTGSRATSGMTVANSFGNGTQVNTVNTALTQSGASV 62

DB 381 GCQVLHGVN-QWNTGFTANTYKNTSSAPVDGWTLTSTFPSCQOVTOAMSSTVTQSGAV 439

QY 63 TATMLSYNNVTOGOSTTTFGNGSYSGTNTAPT 95

DB 440 TVRNAPNNGSIPAGGTAQFGFNGSHTGTNAAPT 472

RESULT 5

GUNA_MICBI STANDARD; PRT; 456 AA.

AC P26414;

DT 01-AUG-1992 (Rel. 23, Created)

DT 01-AUG-1992 (Rel. 23, Last sequence update)

DT 15-DEC-1998 (Rel. 37, Last annotation update)

DE Endoglucanase A precursor (EC 3.2.1.4) (Endo-1,4-beta-glucanase) (Cellulase).

DE CELA.

OS Microbispora bispora.

GN Bacteria; Actinobacteria; Actinobacterla (class); Actinobacterlidae;

OC Actinomycetales; Pseudonocardineae; Pseudonocardaceae;

OC Thermobispora.

OX NCBI_TaxID-2006;

RN [1]

RA Yablonsky M.D., Elliston K.O., Eveleigh D.E.;

RL (In) Coughlan M.P. (eds.);

RL Production, characterization and application of cellulose, hemicellulose and lignin enzyme degrading systems, pp.77-83, Elsevier, London (1989).

CC -1- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic linkages in cellulose.

CC -1- SIMILARITY: CONTAINS 1 BACTERIAL-TYPE CELLULOSE-BINDING DOMAIN (CBD).

CC -1- SIMILARITY: BELONGS TO CELLULOSE FAMILY B (FAMILY 6 OF GLYCOSYL HYDROLASES).

DR HSSP: P26222; ITML.

DR InterPro: IPR001919; Bac_cellose-bind.

DR InterPro: IPR001524; GH_6.

DR Pfam: PF00553; CBM_2; 1.

DR Pfam: PF01341; Glyco_hydro_6; 1.

DR PRINTS: PR00733; GLYDRLASE6.

DR PRODOM: PD003733; GH_6; 1.

DR PROSITE: PS00561; CBD_BACTERIAL; 1.

DR PROSITE: PS00655; GLYCOSYL_HYDROL_F6_1; 1.

DR PROSITE: PS00656; GLYCOSYL_HYDROL_F6_2; 1.

KW Cellulose degradation; Hydrolase; Glycosidase; Signal.

FT SIGNAL 1 30 POTENTIAL.

FT CHAIN 31 456 ENDOGLUCANASE A.

FT DOMAIN 31 322 CATALYTIC.

FT DOMAIN 323 355 LINKER ("HINGE") (PRO-SER BOX).

FT DOMAIN 356 456 CELLULOSE-BINDING.

FT ACT_SITE 113 113 BY SIMILARITY.

FT ACT_SITE 151 151 PROTON DONOR (BY SIMILARITY).

FT ACT_SITE 300 300 NUCLEOPHILE (BY SIMILARITY).

FT DISULFID 114 159 BY SIMILARITY.

FT DISULFID 267 302 BY SIMILARITY.

FT DISULFID 360 453 BY SIMILARITY.

SQ SEQUENCE 456 AA; 47011 MW; B06D8595E322848F CRC64;

Query Match 44.3%; Score 240.5; DB 1; Length 456;

DR	Pfam:	PF00553;	CBM_2; 1.
DR	Pfam:	PF01341;	Glyco_hydro6; 1.
DR	PRINTS:	PR00014;	FNTYPEIIIL
DR	PRINTS:	PD00733;	GLHYDRLASE6.
DR	ProDom:	PD003733;	GH-6; 2.
DR	SMART:	SM00600;	FN3; 3.
DR	PROSITE:	PS00561;	CBD_BACTERIAL; 1.
DR	PROSITE:	PS00655;	GLYCOSYL_HYDROL_F6-1; 1.
DR	PROSITE:	PS00656;	GLYCOSYL_HYDROL_F6-2; 1.
KW	Cellulose degradation; Hydrolase; Glycosidase; Repeat; Signal.		
FT	SIGNAL	1	40
FT	CHAIN	41	872
FT	DOMAIN	41	477
FT	DOMAIN	478	563
FT	DOMAIN	573	664
FT	DOMAIN	673	768
FT	DOMAIN	769	872
FT	ACT_SITE	188	188
FT	ACT_SITE	410	410
FT	DISULFID	140	202
FT	DISULFID	374	428
FT	DISULFID	770	869
SQ	SEQUENCE	872 AA;	89300 MW; 7883B407F995533B CRC64;
Query Match 43.6%; Score 236.5; DB 1; Length 872;			
Best Local Similarity 43.7%; Pred. No. 9e-14;			
Matches 45; Conservative 17; Mismatches 34; Indels 7; Gaps			
Qy	3	GCRATVYVNSDWGSGFTATVTIVTNGSRATSGWTVAHSPCGNQTVTNVNWHTALTOSCASY 62	
Dd	769	GCTVKYASSS-WNFGTGTTVEVKNNGTAALINGWTLGFSADGKGVSGQNSAENWSQGTAV 827	
Qy	63	TATLSLYNNVTPQCSGTFPGNGSYSCNTAPT-----LTCT 99	
Dd	828	TAKNAPWNGTLAAGSSYSIGFNGTHNGTNTAPTFTLNGVACT 870	
RESULT 7			
ID	GUNB_CELFI	STANDARD;	PRT; 1045 AA.
AC	P26225;		
DT	01-MAY-1992 (Rel. 22, Created)		
DT	01-MAY-1992 (Rel. 22, Last sequence update)		
DT	16-OCT-2001 (Rel. 40, Last annotation update)		
DE	Endoglucanase B precursor (EC 3.2.1.4) (Endo-1,4-beta-glucanase B)		
DE	(Cellulase B).		
GN	CenB.		
OC	Cellulomonas fimi.		
OS	Bacteria; Actinobacteria; Actinobacteriales (class); Actinobacteridae;		
OX	Actinomycetales; Micrococciaceae; Cellulomonadaceae; Cellulomonas.		
NCBI_TaxID=1708;			
[1]			
RN	SEQUENCE FROM N.A.		
RX	MEDLINE-91100298; PubMed-1987122;		
RA	Meinke A., Braun C., Gilkes N.R., Kilburn D.G., Miller R.C. Jr.,		
RA	Warren R.A.J.;		
RT	"Unusual sequence organization in CenB, an inverting endoglucanase		
RT	from Cellulomonas fimi."		
J. Bacteriol.	173:308-314(1991).		
RN	[2]		
RP	DOMAINS.		
RX	MEDLINE-92041609; PubMed-1938913;		
RA	Meinke A., Gilkes N.R., Kilburn D.G., Miller R.C. Jr., Warren R.A.J.;		
RT	"Multiple domains in endoglucanase B (CenB) from Cellulomonas fimi;		
RT	functions and relatedness to domains in other polypeptides.";		
RL	J. Bacteriol. 173:7126-7135(1991).		
CC	-1- FUNCTION: THE BIOLOGICAL CONVERSION OF CELLULOSE TO GLUCOSE		
CC	GENERALLY REQUIRES THREE TYPES OF HYDROLYTIC ENZYMES:		
CC	(1) ENDOGLUCANASES WHICH CUT INTERNAL BETA-1,4-GLUCOSIDIC BONDS;		
CC	(2) EXOCELLULOXYDHROLASES THAT CUT THE DISSACCHARIDE CELLOBIOSE		
CC	FROM THE NONREDUCING END OF THE CELLULOSE POLYMER CHAIN;		
CC	(3) BETA-1,4-GLUCOSIDASES WHICH HYDROLYZE THE CELLOBIOSE AND OTHER		
CC	SHORT CELLO-OLIGOSACCHARIDES TO GLUCOSE.		

```

CC CC -!- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic
CC linkages in cellulose.
CC -!- MISCELLANEOUS: THE LINKER REGION (ALSO TERMED "HINGE") MAY BE A
CC POTENTIAL SITE FOR PROTEOLYSIS.
CC -!- MISCELLANEOUS: MAY CONTAIN A SECOND CBD IN THE CATALYTIC DOMAIN.
CC -!- SIMILARITY: CONTAINS 3 FIBRONECTIN TYPE III-LIKE DOMAINS.
CC -!- SIMILARITY: BELONGS TO CELLULOSE FAMILY E (FAMILY 9 OF GLYCOSYL
CC HYDROLASES).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch)
CC -----
CC EMBL: M64644; AAA23086.1; -.
CC PIR: A39199; A39199.
CC HSSP: P26221; ITR4.
CC InterPro: IPR001919; Bac_cellose-bind.
CC InterPro: IPR001956; CBD_3.
CC InterPro: IPR003961; FN.III.
CC InterPro: IPR001701; CH_9.
CC Pfam: PF00441; fn3_3.
CC Pfam: PF00553; CBM_2; 1.
CC Pfam: PF00759; Glyco_hydro_9; 1.
CC Pfam: PF00942; CBM_3; 1.
CC SMART: SM00060; FN3_3.
CC PROSITE: PS00561; CBD_BACTERIAL; 1.
CC PROSITE: PS00592; GLYCOSYL_HYDROL_F9_1; 1.
CC PROSITE: PS00698; GLYCOSYL_HYDROL_F9_2; 1.
CC Cellulose degradation: Hydrolase; Glycosidase; Repeat; Signal.
FT SIGNAL 1 33
FT CHAIN 34 1045
FT DOMAIN 34 643
FT DOMAIN 34 643
FT DOMAIN 644 650
FT DOMAIN 651 733
FT DOMAIN 734 748
FT DOMAIN 749 830
FT DOMAIN 831 846
FT DOMAIN 847 930
FT DOMAIN 931 944
FT DOMAIN 945 1045
FT ACT_SITE 410 410
FT ACT_SITE 449 449
FT ACT_SITE 458 458
FT DISULFID 946 1044
FT SEQUENCE 1045 AA; 108990 MW; AC2F7B84E4E3C4F0 CRC64;

Query Match 43.1%; Score 234; DB 1; Length 1045;
Best Local Similarity 49.5%; Pred. No. 1.8e-13;
Matches 45; Conservative 14; Mismatches 30; Indels 2; Gaps 2;

QY 4 CRATYVNSDMGSGFTATVTNTGSRATSGMTVAWSFGNGQTVTNWNTALTQSGASVT 63
Db 946 CTVVYSTNS-WNVGFTGSKVINTGTTPLT-WTLGEAFPSGGQOVGWSATWSQTGTTVT 1003

QY 64 ATNLNYYNNVQPCQSTTFGNGSYSGTNTAP 94
Db 1004 ATGLSNWNLQPCQSTDIGFNGCHPCTNTNP 1034

RESULT 8
GUNC_PSEFL
ID GUNC_PSEFL STANDARD: PRT: 748 AA.
AC P27033;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 01-JUN-1994 (Rel. 29, Last annotation update)
DE Endoglucanase C precursor (EC 3.2.1.4) (Endo-1,4-beta-glucanase)
DE (Cellulase) (EGC) (Cellodextrinase C).
GN CELC.

```

```

OS Pseudomonas fluorescens.
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
OC Pseudomonas.
OX NCBI_TaxID=294;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 38-47.
RC STRAIN=Sp. Cellulosa;
RX MEDLINE=92061996; PubMed=1953673;
RA Ferreira L.M.A., Hazlewood G.P., Barker P.J., Gilbert H.J.;
RT "The cellodextrinase from Pseudomonas fluorescens subsp. cellulosa
RT consists of multiple functional domains.";
RL Biochem. J. 279:793-799(1991).
CC -!- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic
CC linkages in cellulose.
CC -!- SIMILARITY: CONTAINS 1 BACTERIAL-TYPE CELLULOSE-BINDING DOMAIN
CC (CBD)
CC -!- SIMILARITY: BELONGS TO CELLULOSE FAMILY A (FAMILY 5 OF GLYCOSYL
CC HYDROLASES).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch)
CC -----
CC EMBL: X61299; CAA43597.1; -.
CC PIR: S16849; S16849.
CC HSSP: S19652; S19652.
CC HSSP: P07986; IEXG.
CC InterPro: IPR001919; Bac_cellose-bind.
CC InterPro: IPR002883; CBD_5.
CC InterPro: IPR001547; GH_5.
CC Pfam: PF00150; cellulase; 1.
CC Pfam: PF00553; CBM_2; 1.
CC Pfam: PF02013; CBM_10; 1.
CC PROSITE: PS00561; CBD_BACTERIAL; 1.
CC PROSITE: PS00659; GLYCOSYL_HYDROL_F5; 1.
CC Cellulose degradation: Hydrolase; Glycosidase; Signal.
FT SIGNAL 1 38
FT CHAIN 39 748
FT DOMAIN 39 136
FT DOMAIN 137 179
FT DOMAIN 228 280
FT DOMAIN 281 748
FT DISULFID 39 133
FT ACT_SITE 503 503
FT ACT_SITE 653 653
FT SEQUENCE 748 AA; 80197 MW; AC96104137932B76 CRC64;

Query Match 42.4%; Score 230; DB 1; Length 748;
Best Local Similarity 47.5%; Pred. No. 2.9e-13;
Matches 48; Conservative 16; Mismatches 29; Indels 8; Gaps 5;

QY 3 GCRATYVNSDMGSGFTATVTNTGSRATSGMTVAWSFGNGQTVTNWNTALTQSGASV 62
Db 38 GCE--YVVTNSMGSGPTAAIRITNSTSSVINGNWSQVNSNR-VTNLWNPNLGSGN-PY 93

QY 63 TATNLNYYNNVQPCQSTTFGNG-SYSGTNTAPT---TCT 99
Db 94 SASNLNWSNGTIQPGQTVTFGFGVTSNGTSPVNGAACT 134

RESULT 9
GUNC_THEFU
ID GUNC_THEFU STANDARD: PRT: 880 AA.
AC P26221; Q08167;
DT 01-MAY-1992 (Rel. 22, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Endoglucanase E-4 precursor (EC 3.2.1.4) (Endo-1,4-beta-glucanase E-4)
DE (Cellulase E-4) (Cellulase E4).

```


GN CELD.
OS Thermomonospora fusca.
OC Bacteria; Actinobacteria; Actinobacteriales (class); Actinobacteridae;
OC Actinomycetales; Streptosporangineae; Nocardiopsaceae; Thermobifida.
OX NCBI_TaxID=2021;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=YX;
RX MEDLINE=94028932; PubMed=8215374;
RA Jung E.D., Lao G., Irwin D., Barr B.K., Benjamin A., Wilson D.B.;
RT "DNA sequences and expression in Streptomyces lividans of an
RT exoglucanase gene and an endoglucanase gene from Thermomonospora
RT fusca.";
RL Appl. Environ. Microbiol. 59:3032-3043(1993).
RN [2]
RP REVISIONS.
RA Wilson D.B.;
RL Submitted (FEB-1997) to the EMBL/GenBank/DBJ databases.
RN [3]
RP PARTIAL SEQUENCE FROM N.A.
RC STRAIN=YX;
RX MEDLINE=91258320; PubMed=1904434;
RA Lao G., Ghangas G.S., Jung E.D., Wilson D.B.;
RT "DNA sequences of three beta-1,4-endoglucanase genes from
RT Thermomonospora fusca.";
RL J. Bacteriol. 173:3397-3407(1991).
RN [4]
RP SEQUENCE OF 47-67.
RA Wilson D.B.;
RL "Cellulases of Thermomonospora fusca.";
RT Meth. Enzymol. 160:314-323(1988).
RN [5]
RP X-RAY CRYSTALLOGRAPHY (1.9 ANGSTROMS) OF 47-651.
RX MEDLINE=97475222; PubMed=9334746;
RA Sakon J., Irwin D., Wilson D.B., Karplus P.A.;
RT "Structure and mechanism of endo/exocellulase E4 from Thermomonospora
RT fusca.";
RL Nat. Struct. Biol. 4:810-818(1997).
CC -1- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic
CC linkages in cellulose.
CC -1- PATHWAY: Cellulose degradation.
CC -1- SIMILARITY: BELONGS TO CELLULASE FAMILY E (FAMILY 9 OF GLYCOSYL
CC HYDROLASES).
CC -1- SIMILARITY: CONTAINS 1 FUNGAL-TYPE CELLULOSE-BINDING DOMAIN (CBD).
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: L20093; AAB42155.1;
DR EMBL: M73322; AAA27397.1; ALT_SEQ.
DR IIR; B42360; B42360.
DR PDB: 1J54; 17-SEP-97.
DR PDB: 1TF4; 04-SEP-97.
DR PDB: 3TF4; 04-SEP-97.
DR PDB: 4TF4; 04-SEP-97.
DR InterPro: IPR001919; Bac_cellose-bind.
DR InterPro: IPR001956; CBD_3.
DR InterPro: IPR003961; FN_III.
DR InterPro: IPR003962; FNIII_repeat.
DR InterPro: IPR001701; GH_9.
DR InterPro: IPR001230; Prenyl_site.
DR Pfam: PF00041; fn3; 1.
DR Pfam: PF00553; CBM_2; 1.
DR Pfam: PF00759; Glyco_hydro_9; 1.
DR Pfam: PF00942; CBM_3; 1.
DR PRINTS; PR00014; FNTYPEIII.
DR SMART; SM00060; FN3; 1.
DR PROSITE; PS00561; CBD_BACTERIAL; 1.

DR PROSITE; PS00592; GLYCOSYL_HYDROL_F9_1; 1.
DR PROSITE; PS00698; GLYCOSYL_HYDROL_F9_2; 1.
KW Cellulose degradation; Hydrolase; Glycosidase; Signal; 3D-structure.
FT SIGNAL 1 46
FT CHAIN 47 880 ENDOGLUCANASE E-4.
FT DOMAIN 776 880 CELLULOSE-BINDING.
FT ACT_SITE 427 427 BY SIMILARITY.
FT ACT_SITE 461 461 BY SIMILARITY.
FT ACT_SITE 470 470 BY SIMILARITY.
SQ SEQUENCE 880 AA; 95202 MW; 5EA9A6ABF45A4D9A CRC64;
Query Match 41.5%; Score 225.5; DB 1; Length 880;
Best Local Similarity 45.7%; Pred. No. 8.6e-13;
Matches 42; Conservative 17; Mismatches 32; Indels 1; Cuts 1;
QY 4 CRATVYVNDGSGFTATVTVTNGSRATSGTGTVMVSGGNOTVTNYNNTALTQSGASYT 63
DB 778 CTGVYSTN-DWDSGFTASIRITYHCTAPLSSWELSFYTPAGQOVTHGNATWRDCAAVT 836
QY 64 ATNLSYNNVIOPGOSTTTFGNGSYSGTNTAPT 95
DB 837 ATPMSWNSLAPGATVEVGFNGSWGSGNTPTPT 868
RESULT 10
GUXB_CELFI STANDARD: PRT; 1090 AA.
ID GUXB_CELFI
AC PS0899;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE Exoglucanase B precursor (EC 3.2.1.91) (Exocellobiohydrolase B)
DE (1,4-beta-cellobiohydrolase B) (CBP120).
GN CBHB OR CENE.
OS Cellulomonas fimi.
OC Bacteria; Actinobacteria; Actinobacteriales (class); Actinobacteridae;
OC Actinomycetales; Micrococccineae; Cellulomonadaceae; Cellulomonas.
OX NCBI_TaxID=1708;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 456-461.
RC STRAIN=ATCC 484;
RX MEDLINE=96003898; PubMed=7575482;
RA Shen H., Gilkes N.R., Kilburn D.G., Miller R.C. Jr., Warren R.A.J.;
RT "Cellulobiohydrolase B, a second exo-cellobiohydrolase from the
RL cellulolytic bacterium Cellulomonas fimi.";
RN Biochem. J. 311:67-74(1995).
RP [2]
RP SEQUENCE OF 54-75.
RX MEDLINE=93209933; PubMed=8458833;
RA Meinke A., Gilkes N.R., Kilburn D.G., Miller R.C. Jr., Warren R.A.J.;
RT "Cellulose-binding polypeptides from Cellulomonas fimi: endoglucanase
RL D (Cend), a family A beta-1,4-glucanase.";
RL J. Bacteriol. 175:1910-1918(1993).
RN [3]
RP SEQUENCE OF 54-78.
RX MEDLINE=94197708; PubMed=8147863;
RA Shen H., Tomme P., Meinke A., Gilkes N.R., Kilburn D.G.,
RA Warren R.A.J., Miller R.C. Jr.;
RT "Stereochemical course of hydrolysis catalysed by Cellulomonas fimi
RL Cene, a member of a new family of beta-1,4-glucanases.";
RL Biochem. Biophys. Res. Commun. 199:1223-1228(1994).
CC -1- FUNCTION: HYDROLYSE CELLOHEXAOSE TO A MIXTURE OF CELLOTETRAOSE,
CC CELLOTRIOSE AND CELLOBIOSE, WITH ONLY A TRACE OF GLUCOSE. IT
CC HYDROLYSED CELLOPENTAOSE TO CELLOTRIOSE AND CELLOBIOSE, AND
CC CELLOTETRAOSE TO CELLOBIOSE, BUT IT DID NOT HYDROLYSE CELLOTRIOSE.
CC HAS ALSO WEAK ENDOGLUCANASE ACTIVITY. HYDROLYSES GLUCOSIDIC BONDS
CC WITH INVERSION OF ANOMERIC CONFIGURATION.
CC -1- CATALYTIC ACTIVITY: Hydrolysis of 1,4-beta-D-glucosidic linkages
CC in cellulose and cellotetraose, releasing cellobiose from the non-
CC reducing ends of the chains.
CC -1- SIMILARITY: CONTAINS 3 FIBRONECTIN TYPE III-LIKE DOMAINS.
CC -1- SIMILARITY: CONTAINS 1 BACTERIAL-TYPE CELLULOSE-BINDING DOMAIN
CC (CBD).

```
CC -!- SIMILARITY: BELONGS TO CELLULOSE FAMILY L (FAMILY 48 OF GLYCOSYL
CC HYDROLASES).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: L38827; AAB00822.1; -.
CC HSSP: P07986; 1EXG.
CC InterPro: IPR001919; Bac_cellose-bind.
CC InterPro: IPR003961; FN.III.
CC InterPro: IPR003962; FN.III-repeat.
CC InterPro: IPR000556; Glyco_hydro_48.
CC Pfam: PF00041; fn3; 3.
CC Pfam: PF00553; CBM_2; 1.
CC PRINTS: PR02011; Glyco_hydro_48; 1.
CC PRINTS: PR00014; FNTYPEIIL.
CC PRINTS: PR00844; GLHYDRLASE48.
CC PRODOM: PD011903; Glyco_hydro_48; 1.
CC SMART: SM00060; FN3; 3.
CC PROSITE: PS00561; CBD_BACTERIAL; 1.
CC Cellulose degradation: Hydrolase; Glycosidase; Repeat; Signal.
CC SIGNAL 1 33
CC PROPEP 34 53
CC CHAIN 54 1090
CC DOMAIN 54 699
CC DOMAIN 700 785
CC DOMAIN 794 884
CC DOMAIN 891 978
CC DOMAIN 989 1090
CC ACT_SITE 513 513
CC DISULFID 990 1089
CC SEQUENCE 1090 AA; 114829 MW; 046BB9D956F2F399 CRC64;
Query Match 41.2%; Score 223.5; DB 1; Length 1090;
Best Local Similarity 41.0%; Pred. No. 1.6e-12;
Matches 43; Conservative 17; Mismatches 38; Indels 7; Caps 2;
QY 1 GVCGRATYVNVDSGSGFTATVTVTNGSRATSGWTVAMSGNGQTVNTVNTALQTQSGA 60
DB 987 GSGCAYNAYNASS-WNSGFTASVITATGTTTNGSLGDLTACQKVOQGSATWTQSGS 1045
QY 61 SVTATNLNYYNVIQPGSTTFGNGSYSGTNAFT-----LTCT 99
DB 1046 TVTATNAPNNGTLAPQCTVDVGVNGSHGQNPASFTLNGASCT 1090
RESULT 11
CHIT_STRLI
ID CHIT_STRLI STANDARD; PRT; 619 AA.
AC P3609;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DE 01-FEB-1995 (Rel. 31, Last annotation update)
DE Chitinase C precursor (EC 3.2.1.14).
GN CHIC.
OS Streptomyces lividans.
OC Bacteria; Actinobacteria; Actinobacteria (class); Actinobacteridae;
OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1916;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93294525; PubMed=8515228;
RA Fujii T., Miyashita K.;
RT "Multiple domain structure in a chitinase gene (chic) of Streptomyces
RL J. Gen. Microbiol. 139:677-686(1993).
CC -!- CATALYTIC ACTIVITY: Hydrolysis of the 1,4-beta-linkages of N-
acetyl-D-glucosamine polymers of chitin.
```

```
CC -!- INDUCTION: BY CHITIN.
CC -!- SIMILARITY: BELONGS TO CHITINASE CLASS 11 (FAMILY 18 OF GLYCOSYL
CC HYDROLASES).
CC -!- SIMILARITY: CONTAINS 1 FIBRONECTIN TYPE III-LIKE DOMAIN.
CC -!- SIMILARITY: CONTAINS 1 BACTERIAL-TYPE CELLULOSE-BINDING DOMAIN
CC (CBD).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: D12647; BAA02168.1; -.
CC HSSP: P07986; 1EXG.
CC InterPro: IPR001919; Bac_cellose-bind.
CC InterPro: IPR001579; Chitinase_18/2.
CC InterPro: IPR002048; EF-hand.
CC InterPro: IPR003961; FN.III.
CC InterPro: IPR001223; Glyco_hydro_18.
CC Pfam: PF00041; fn3; 1.
CC Pfam: PF00553; CBM_2; 1.
CC PRODOM: PD000471; Glyco_hydro_18; 1.
CC SMART: SM00060; FN3; 1.
CC PROSITE: PS00018; EF_HAND; UNKNOWN_1.
CC PROSITE: PS00561; CBD_BACTERIAL; 1.
CC PROSITE: PS01095; CHITINASE_18; 1.
CC Hydrolase; Glycosidase; Chitin degradation; Chitin-binding; Signal.
CC SIGNAL 1 30
CC CHAIN 31 619
CC DOMAIN 35 140
CC DOMAIN 148 230
CC DOMAIN 240 619
CC ACT_SITE 382 382
CC SEQUENCE 619 AA; 65200 MW; A23CEE5B3C5D6F21 CRC64;
Query Match 39.2%; Score 213; DB 1; Length 619;
Best Local Similarity 45.9%; Pred. No. 7.6e-12;
Matches 39; Conservative 13; Mismatches 33; Indels 0; Caps 0;
QY 6 ATTVVNSDMSGSGFTATVTVTNGSRATSGWTVAMSGNGQTVNTVNTALQTQSGAVTAT 65
DB 36 ATAKYSDMGTFGSGSWTVANTGTTSLSSMTVEHDFPTGTYKTSAMDATVTVNSGDHWTAK 95
QY 66 NLSYNNVIQPGSTTFGNGSYSGT 90
DB 96 NVGWNGLAPGASVSGFGNGSGPCS 120
RESULT 12
GUNA_CELFI
ID GUNA_CELFI STANDARD; PRT; 449 AA.
AC P07984;
DT 01-AUG-1988 (Rel. 08, Created)
DT 01-AUG-1988 (Rel. 08, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE Endoglucanase A precursor (EC 3.2.1.4) (Endo-1,4-beta-glucanase)
DE (Cellulase).
GN CENA.
OS Cellulomonas fimi.
OC Bacteria; Actinobacteria; Actinobacteria (class); Actinobacteridae;
OC Actinomycetales; Micrococineae; Cellulomonadaceae; Cellulomonas.
OX NCBI_TaxID=1708;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=87055249; PubMed=3023193;
RA Wong W.K.R., Gerhard B., Guo Z.M., Kilburn D.G., Warren R.A.J.,
RA Miller R.C. Jr.;
RT "Characterization and structure of an endoglucanase gene cenA of
Cellulomonas fimi.";
```


QY	67	LSYNNVLIQP-QGSTTFGFGNGSYSGTNTAPT 95
		: : :
DB	474	AGYNGTIAANGQTQSFGFNINSGVLKPT 503
		: :
 RESULT 14		
ID	CHIT_STRPL	STANDARD; PRT: 610 AA.
AC	P11220;	
DT	01-JUL-1989 (Rel. 11, Created)	
DT	01-DEC-1992 (Rel. 24, Last sequence update)	
DT	01-FEB-1995 (Rel. 31, Last annotation update)	
DE	Chitinase 63 precursor (EC 3.2.1.14).	
GN	CHTA.	
OS	Streptomyces plicatus.	
OC	Bacteria; Actinobacteria; Actinobacteriales (class); Actinobacteridae;	
OC	Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.	
OX	NCBI_TaxID=1922;	
RN	[1]	
SEQUENCE FROM N.A.		
RX	MEDLINE=92192480; PubMed=1532161;	
RY	Robbins P.W., Overbye K., Albright C., Benfield B., Pero J.;	
RT	"Cloning and high-level expression of chitinase-encoding gene of	
RT	Streptomyces plicatus";	
RL	Gene 111:69-76(1992).	
	[2]	
RN	SEQUENCE OF 1-45 FROM N.A., AND SEQUENCE OF 31-45.	
RX	MEDLINE=88087127; PubMed=3275646;	
RY	Robbins P.W., Albright C., Benfield B.;	
RT	"Cloning and expression of a Streptomyces plicatus chitinase	
RT	(chitinase-63) in Escherichia coli.";	
J. Biol. Chem.	263:443-447(1988).	
CC	-I- CATALYTIC ACTIVITY: Hydrolysis of the 1,4-beta-linkages of N-	
CC	acetyl-D-glucosamine polymers of chitin.	
CC	-I- INDUCTION: BY CHITIN.	
CC	-I- SIMILARITY: BELONGS TO CHITINASE CLASS II (FAMILY 18 OF GLYCOSYL	
CC	HYDROLASES).	
CC	-I- SIMILARITY: CONTAINS 1 FIBRONECTIN TYPE III-LIKE DOMAIN.	
CC	-I- SIMILARITY: CONTAINS 1 BACTERIAL-TYPE CELLULOSE-BINDING DOMAIN	
	(CBD).	

This SWISS-PROT entry is copyright. It is produced through a collaboration		
between the Swiss Institute of Bioinformatics and the EMBL Outstation -		
the European Bioinformatics Institute. There are no restrictions on its		
use by non-profit institutions as long as its content is in no way		
modified and this statement is not removed. Usage by and for commercial		
entities requires a license agreement (see http://www.isb-sib.ch/announce/		
or send an email to license@isb-sib.ch).		

EMBL:	M82804; AAA26720.1; -;	
DR	EMBL: M18397; AAA26717.1; -;	
DR	PIR: A29912; A29912.	
DR	PIR: JH0573; JH0573.	
DR	HSSP: P07986; IEXG.	
DR	InterPro: IPR001919; Bac_celose-bind.	
DR	InterPro: IPR001579; Chitinase_18/2.	
DR	InterPro: IPR002048; EF-hand.	
DR	InterPro: IPR003961; FN_III.	
DR	InterPro: IPR001223; Glyco_hydro_18.	
DR	Pfam: PF00041; fn3; 1.	
DR	Pfam: PF00553; CBM_2; 1.	
DR	Pfam: PF00704; Glyco_hydro_18; 1.	
DR	ProDom: PD000471; Glyco_hydro_18; 1.	
DR	SMART: SM00060; FN3; 1.	
DR	PROSITE: PS00018; EF_HAND; UNKNOWN_1.	
DR	PROSITE: PS00561; CBD_BACTERIAL; 1.	
DR	PROSITE: PS01095; CHITINASE_18; 1.	
KW	Hydrolase; Glycosidase; Chitin degradation; Chitin-binding; Signal.	
FT	SIGNAL	1 30
FT	CHAIN	31 610 CHITINASE 63.
FT	DOMAIN	35 140 CELLULOSE-BINDING.
FT	DOMAIN	148 229 FIBRONECTIN TYPE-III.

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 13, 2002, 10:50:41 : Search time 9.35991 Seconds
(without alignments)
2223.392 Million cell updates/sec

Title: US-09-917-378-5

Perfect score: 543

Sequence: 1 GVGCRTYVVNSDMGSGFTA.....GFNGSYSGTNTAPTLTCTAS 101

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 20000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL_21:*

- 1: sp_archaea:*
- 2: sp_bacteria:*
- 3: sp_fungi:*
- 4: sp_human:*
- 5: sp_invertebrate:*
- 6: sp_mammal:*
- 7: sp_mhc:*
- 8: sp_organelle:*
- 9: sp_phase:*
- 10: sp_plant:*
- 11: sp_rodent:*
- 12: sp_virus:*
- 13: sp_vertebrate:*
- 14: sp_unclassified:*
- 15: sp_rviro:*
- 16: sp_bacteriap:*
- 17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	292.5	53.9	579	16	O86730 streptomyc
2	275	50.6	596	2	Q60029 thermomonos
3	265	48.8	984	2	Q9XCD4
4	257	47.3	890	16	O86727 streptomyc
5	254	46.8	495	2	O52799 amycolatops
6	247.5	45.6	881	2	O9AF09 pseudomonas
7	244.5	45.0	485	2	O59277 cellulomona
8	235	43.3	623	2	O8ROP6 streptomyc
9	228.5	42.1	381	2	O54331
10	228.5	42.1	381	16	O9RJY3 streptomyc
11	228.5	42.1	973	16	O86728 streptomyc
12	224	41.3	629	2	O9L8G0 streptomyc
13	220	40.5	974	2	O08166 thermomonos
14	213	39.2	270	2	O9S675 streptomyc
15	213	39.2	609	16	Q9Z9M8 streptomyc
16	204.5	37.7	377	2	O08468 streptomyc

17	198	36.5	570	2	Q59665	Q59665 pseudomonas
18	197	36.3	400	16	O69962	O69962 streptomyc
19	197	36.3	649	2	O9F7L3	O9F7L3 pseudomonas
20	195.5	36.0	450	16	O9RJY1	O9RJY1 streptomyc
21	192.5	35.5	482	2	P74912	P74912 thermobifid
22	191.5	35.3	382	2	O59963	O59963 streptomyc
23	189	34.8	364	16	O9RJY2	O9RJY2 streptomyc
24	187.5	34.5	364	2	O07862	O07862 streptomyc
25	187	34.4	486	16	O9F3Q0	O9F3Q0 streptomyc
26	186	34.3	358	2	O9X5K8	O9X5K8 streptomyc
27	184.5	34.0	583	2	O51815	O51815 pseudomonas
28	178.5	32.9	457	2	O53488	O53488 micromonosp
29	162.5	29.9	371	2	O9KIH1	O9KIH1 streptomyc
30	161.5	29.7	461	2	O59922	O59922 streptomyc
31	156	28.7	571	2	O9EY44	O9EY44 cellulomona
32	144	26.5	745	2	O9K5C7	O9K5C7 clavibacter
33	137	25.2	683	16	O9PDW2	O9PDW2 xyliella fas
34	134.5	24.8	335	2	O9RQB8	O9RQB8 cellulomona
35	133.5	24.6	635	2	O66065	O66065 fibrobacter
36	132.5	24.4	592	16	O9PF60	O9PF60 xyliella fas
37	131	24.1	674	2	O9FAC8	O9FAC8 bacillus ce
38	129	23.8	635	2	O07088	O07088 bacillus th
39	125	23.0	481	2	O66064	O66064 actinomycas
40	125	23.0	727	2	O9AF65	O9AF65 clavibacter
41	123.5	22.7	384	2	O9X602	O9X602 streptomyc
42	123	22.7	676	2	O93AD3	O93AD3 bacillus th
43	121	22.3	1070	16	O55365	O55365 synethocyst
44	118.5	21.8	489	16	O9F312	O9F312 streptomyc
45	116.5	21.5	765	16	O54183	O54183 streptomyc

ALIGNMENTS

RESULT 1

O86730

ID O86730 PRELIMINARY: PRT: 579 AA.

AC O86730:

DT 01-NOV-1998 (TREMBLrel. 08, Created)

DT 01-NOV-1998 (TREMBLrel. 08, Last sequence update)

DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)

DE Putative secreted cellulase.

GN SC06548 OR SC5C7.33.

OS Streptomyces coelicolor.

OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;

OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomycetes.

OX NCBI_TaxID=1902;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=A3(2);

RA Seeger K.J., Harris D.;

RL Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.

RN [2]

RP SEQUENCE FROM N.A.

RC STRAIN=A3(2);

RA Parkhill J., Barrell B.G., Rajandream M.A.;

RL Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.

RN [3]

RP SEQUENCE FROM N.A.

RC STRAIN=A3(2);

RX MEDLINE=97000351; PubMed=8843436;

RA Redenbach M., Kieser H.M., Denapalte D., Eichner A., Cullum J.,

RA Kinashi H., Hopwood D.A.;

RT "A set of ordered cosmids and a detailed genetic and physical map for the 8 Mb Streptomyces coelicolor A3(2) chromosome.";

RL Mol. Microbiol. 21:77-96(1996).

RN [4]

RP SEQUENCE FROM N.A.

RC STRAIN=A3(2) / M145;

RA Bentley S.D., Chater K.F., Cerdeno-Tarraga A.-M., Challis G.L.,

RA Thomson N.R., James K.D., Harris D.E., Quail M.A., Kieser H.,

RA Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,

RA Cronin A., Fraser A., Goble A., Hidalgo J., Hornsby T., Howarth S.,

```

RA Huang C.-H., Kieser T., Larke L., Murphy L., Oliver K., O'Neil S.,
RA Rabinowitz E., Rajandream M.A., Rutherford K., Rutter S.,
RA Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K.,
RA Warren T., Wietzorrek A., Woodward J., Barrell B.G., Parkhill J.,
RA Hopwood D.A.;
RT "Complete genome sequence of the model actinomycete Streptomyces
RT coelicolor A3(2).";
RL Nature 417:141-147(2002).
DR EMBL: AL031515; CAA20645.1; -.
DR HSP: P07986; IEXG.
DR InterPro: IPR001919; Bac_cellose-bind.
DR InterPro: IPR001524; GH_6.
DR Pfam: PF00553; CBM_2; 1.
DR Pfam: PF01341; Glyco_hydro_6; 1.
DR PRINTS: PR00733; GLHYDRLASE6.
DR PRODOM: PD003733; GH_6; 1.
DR PROSITE: PS00655; GLYCOSYL_HYDROL_F6_1; 1.
SQ SEQUENCE 579 AA: 60831 MW: 5A51DD28E7996A89 CRC64;

Query Match 53.9%; Score 292.5; DB 16; Length 579;
Best Local Similarity 53.5%; Pred. No. 3.4e-16;
Matches 54; Conservative 14; Mismatches 26; Indels 7; Gaps 2;

QY 4 CRATYVNSDWSGFTATVTVTNGSRATSGWTVAWSFGNQTVTYNTALTQSGASVT 63
DB 37 CTVDYQVQNDWGSCTAAVTVTNNGA-ATSSWSLGTVAGSKVTNSNAKVTSQGAAT 95

QY 64 ATNLSYNNVIOPGSTTFGFGNSYSGTNTAPT-----LTC 98
DB 96 AANESYNGTLTGSGASFGFGTQVSGSNAIPATFTLNGVTC 136

RESULT 2
Q60029 PRELIMINARY; PRT: 596 AA.
ID Q60029
AC Q60029;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Beta-1,4-exocellulase precursor (EC 3.2.1.91).
CN E3.
OS Thermomonospora fusca.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptosporangineae; Nocardiopsaceae; Thermobifida.
OX NCBI_TaxID=2021;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=YX;
RA Zhang S., Lao G., Wilson D.B.;
RT "Characterization of a Thermomonospora fusca exocellulase.";
RL Biochemistry 34:0-0(1995).
DR EMBL: U18978; AAA62211.1; -.
DR HSP: P07986; IEXG.
DR InterPro: IPR001919; Bac_cellose-bind.
DR InterPro: IPR001524; GH_6.
DR Pfam: PF00553; CBM_2; 1.
DR Pfam: PF01341; Glyco_hydro_6; 1.
DR PRINTS: PR00733; GLHYDRLASE6.
DR PRODOM: PD003733; GH_6; 2.
DR PROSITE: PS00655; GLYCOSYL_HYDROL_F6_1; 1.
KW Glycosidase; Hydrolase; Signal.
FT SIGNAL 1 38
FT CHAIN 39 596
FT SEQUENCE 596 AA: 63547 MW: B0FA5277FE7721E0 CRC64;

Query Match 50.6%; Score 275; DB 2; Length 596;
Best Local Similarity 52.4%; Pred. No. 9.2e-15;
Matches 55; Conservative 10; Mismatches 32; Indels 8; Gaps 3;

QY 3 CRATYVNSDWSGFTATVTVTNGSRATSGWTVAWSFGNQTVTYNTALTQSGASV 62
DB 40 GCSVDYTVNS-WGFGFTANVTITNLGS-AINGWTLWDFFGCGNQVTLNLTNGTYTQSGQHV 97

```

```

QY 63 TATNLSYNNVIOPGSTTFGFGNSYSGTNTAPT-----LTCTAS 101
DB 98 SVSNAPYNASIPANGTVEFGFGNSYSGSNDIPSSFKLNCVTCDS 142

RESULT 3
Q9XCD4 PRELIMINARY; PRT: 984 AA.
ID Q9XCD4
AC Q9XCD4;
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Beta-1,4-exocellulase E6 precursor.
CN CELF.
OS Thermomonospora fusca.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptosporangineae; Nocardiopsaceae; Thermobifida.
OX NCBI_TaxID=2021;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=YX;
RA Irwin D.I., Zhang S., Wilson D.B.;
RT "Characterization of a Thermomonospora fusca family 48 exocellulase E6.";
RL Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF144563; AAD39947.1; -.
DR HSP: P07986; IEXG.
DR InterPro: IPR001919; Bac_cellose-bind.
DR InterPro: IPR000556; Glyco_hydro_48.
DR InterPro: IPR000601; PKD_domain.
DR Pfam: PF00553; CBM_2; 1.
DR Pfam: PF02011; Glyco_hydro_48; 1.
DR Pfam: PF00801; PKD; 1.
DR PRINTS: PR00844; GLHYDRLASE48.
DR PRODOM: PD011903; Glyco_hydro_48; 1.
DR SMART: SM00089; PKD; 1.
KW SIGNAL.
FT SIGNAL 1 34
FT CHAIN 35 984
FT SEQUENCE 984 AA: 107127 MW: 9F921C5756AC6153 CRC64;

Query Match 48.8%; Score 265; DB 2; Length 984;
Best Local Similarity 48.1%; Pred. No. 1e-13;
Matches 51; Conservative 13; Mismatches 36; Indels 6; Gaps 1;

QY 2 VCGRATYVNSDWSGFTATVTVTNGSRATSGWTVAWSFGNQTVTYNTALTQSGAS 61
DB 36 VACSVDYDDSDNDWGSFGFVAEVKVTNEGSDPIQNVQVGTFFGCGNQITNGWGVFSOSGAN 95

QY 62 VTATNLSYNNVIOPGSTTFGFGNSYSGTNTAPT-----LTCTAS 101
DB 96 VTVRYPDWNPNTAPGATISFGFGTSGSNDAPTSFTVNGVTCGSGS 141

RESULT 4
Q86727 PRELIMINARY; PRT: 890 AA.
ID Q86727
AC Q86727;
DT 01-NOV-1998 (TrEMBLrel. 08, Created)
DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Putative secreted cellulase.
GN SC06545 OR SC5C7.30C.
OS Streptomyces coelicolor.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1902;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RA Seeger K.J., Harris D.;
RL Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.
RN [2]

```



```

RP SEQUENCE FROM N.A.
RC STRAIN-A3(2);
RA Parkhill J., Barrell B.G., Rajandream M.A.;
RL Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-A3(2);
RX MEDLINE=97000351; PubMed=8843436;
RA Redenbach M., Kleser H.M., Denapalte D., Eichner A., Cullum J.,
RA Kinashi H., Hopwood D.A.;
RT "A set of ordered cosmids and a detailed genetic and physical map for
RL the 8 Mb Streptomyces coelicolor A3(2) chromosome.";
RN Mol. Microbiol. 21:77-96(1996).
[4]
RP SEQUENCE FROM N.A.
RC STRAIN-A3(2) / M145;
RA Bentley S.D., Chater K.F., Cerdeno-Tarraga A.-M., Challis G.L.,
RA Thomson N.R., James K.D., Harris D.E., Quail M.A., Kleser H.,
RA Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,
RA Cronin A., Fraser A., Goble A., Hidalgo J., Hornsby T., Howarth S.,
RA Huang C.-H., Kleser T., Larke L., Murphy L., Oliver K., O'Neill S.,
RA Rabinowitsch E., Rajandream M.A., Rutherford K., Rutter S.,
RA Seeger K., Saunders D., Sharp S., Squares R., Taylor K.,
RA Warten T., Wietzorrek A., Woodward J., Barrell B.G., Parkhill J.,
RA Hopwood D.A.;
RT "Complete genome sequence of the model actinomycete Streptomyces
RL coelicolor A3(2).";
RN Nature 417:141-147(2002).
DR EMBL: AL031515; CAA20642.1;
DR HSSP: P07986; 1EXG.
DR InterPro: IPR001919; Bac_cellose-bind.
DR InterPro: IPR002860; GH_BNR.
DR Pfam: PF02012; BNR; 10.
DR Pfam: PF00553; CBM_2; 1.
SQ SEQUENCE 890 AA; 93252 MW; D2C58695A4B56E84 CRC64;

Query Match 47.3%; Score 257; DB 16; Length 890;
Best Local Similarity 51.6%; Pred. No. 4; 1e-13;
Matches 47; Conservative 10; Mismatches 34; Indels 0; Gaps 0;

QY 4 CRATYVNSDWSGSGFTATVTNTGSRATSGWTVAMSGNQTVNTYNTALTOSGASVT 63
DB 786 CEVTYTVTNQWPGFQADVRLTNTGTSAMNGNSLDWSFGQGVTRMNAERTOAGTSVT 845
QY 64 ATNLSYNNVTPQCGSTTFGNGSYSGTNTAP 94
DB 846 ARNVGNAGVAPGASVGFGTGSRSGTNAEP 876

RESULT 5
O52799 PRELIMINARY; PRT; 495 AA.
AC O52799;
DT 01-JUN-1998 (TrEMBLrel. 06, Created)
DT 01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE PCZA361.11.
OS Anycolatopsis orientalis.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Pseudonocardiaceae; Pseudonocardiaceae; Anycolatopsis.
OX NCBI_TaxID=31958;
RN [1]
RP SEQUENCE FROM N.A.
RA van Wageningen A., Kirkpatrick P., Williams D., Harris B., Kershaw J.,
RA Lennard N., Jones M., Jones S., Solenberg P.;
RT "Sequencing and analysis of genes involved in the biosynthesis of a
RT vancomycin group antibiotic.";
RL Chem. Biol. 3:155-162(1998).
DR EMBL: AJ223998; CAA11769.1;
DR HSSP: P07986; 1EXG.
DR InterPro: IPR001919; Bac_cellose-bind.
DR Pfam: PF00553; CBM_2; 1.
SQ SEQUENCE 495 AA; 52516 MW; 8F54AF6C4761662A CRC64;

```

```

Query Match 46.8%; Score 254; DB 2; Length 495;
Best Local Similarity 49.5%; Pred. No. 3; 9e-13;
Matches 46; Conservative 16; Mismatches 31; Indels 0; Gaps 0;

QY 2 VGRATYVNSDWSGSGFTATVTNTGSRATSGWTVAMSGNQTVNTYNTALTOSGAS 61
DB 392 VDCSVAYKVDQWQNGFTASVTNTGCTTAIDGSLRWTFFENSOOVSNWNAAVRQSDQA 451
QY 62 VTATNLSYNNVTPQCGSTTFGNGSYSGTNTAP 94
DB 452 VSAANWNTKIDPKGSVTFGLASHSLNPPP 484

RESULT 6
O9AF09 PRELIMINARY; PRT; 881 AA.
AC O9AF09;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Rhamnogalacturonan lyase.
OS Pseudomonas cellulosa.
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
OC Pseudomonas.
OX NCBI_TaxID=155077;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 55703;
RX MEDLINE=21157041; PubMed=11256961;
RA McKie V.A., Vincken J.P., Voragen A.G.J., Van Den Broek L.A.M.,
RA Stimson E., Gilbert H.J.;
RT "A new family of rhamnogalacturonan lyases contains an enzyme that
RT binds to cellulose.";
RN Biochem. J. 353:167-177(2001).
CC 1- SIMILARITY: CONTAINS 1 FIBRONECTIN TYPE III-LIKE DOMAIN.
DR EMBL: AY026755; AAK20911.1;
DR HSSP: P07986; 1EXG.
DR InterPro: IPR001919; Bac_cellose-bind.
DR InterPro: IPR003961; FN_III.
DR Pfam: PF00553; CBM_2; 1.
DR Pfam: PF00041; fn3; 1.
DR SMART: SM00060; FN3; 1.
DR PROSITE: PS00561; CBD_BACTERIAL; 1.
KW Hydrolase; Lyase.
SQ SEQUENCE 881 AA; 93963 MW; E8792769751FEEC7 CRC64;

Query Match 45.6%; Score 247.5; DB 2; Length 881;
Best Local Similarity 45.7%; Pred. No. 2; 4e-12;
Matches 42; Conservative 21; Mismatches 26; Indels 3; Gaps 2;

QY 7 TYVNSDWSGSGFTATVTNTGSRATSGWTVAMSGNQTVNTYNTALTOSGAS-VTAT 65
DB 787 SYVITNMGSGFTCAIRITNRGSSAINGWNVWTSYSGNTRISNSWNA--TVSGSNPYSAA 844
QY 66 NLSYNNVTPQCGSTTFGNGSYSGTNTAPTLT 97
DB 845 NLGNATIQGVTFQFGGTGSGTETPVIS 876

RESULT 7
O59277 PRELIMINARY; PRT; 485 AA.
AC O59277;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Exo-beta-1,4-glucanase precursor (EC 3.2.1.91).
GN CEX.
OS Cellulomonas fimi.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Micrococciaceae; Cellulomonadaceae; Cellulomonas.
OX NCBI_TaxID=1708;

```

```
RN
RP
RX MEDLINE=87055250; PubMed=3096818;
RA O'Neill G., Goh S.H., Warren R.A., Kilburn D.G., Miller R.C. Jr.;
RT "Structure of the gene encoding the exoglucanase of Cellulomonas
RL fimi.";
RL Gene 44:325-330(1986).
RN
RP
RP SEQUENCE FROM N.A.
RA Ramirez C., MacLeod A.M., Gilkes N.R.;
RT "Revision to the Cellulomonas fimi Cex gene sequence encoding an exo-
RT beta-1,4-glucanase and beta-1,4-xylanase.";
RL Submitted (FEB-1993) to the EMBL/GenBank/DBJ databases.
DR EMBL; L11080; AAA56792.1; -.
DR HSSP; P07986; 1EXP.
DR InterPro; IPR001919; Bac_cellose-bind.
DR InterPro; IPR001000; Glyco_hydro_10.
DR InterPro; IPR001230; Prenyl_site.
DR Pfam; PF00553; CBM_2; 1.
DR Pfam; PF00331; Glyco_hydro_10; 1.
DR PRINTS; PR00134; GLHYDRASE10.
DR PROSITE; PS00561; CBD_BACTERIAL; 1.
DR PROSITE; PS00591; GLYCOSYL_HYDROL_F10; 1.
DR PROSITE; PS00294; PRENYLATION; UNKNOWN_1.
KW Glycosidase; Hydrolase; Signal.
FT SIGNAL 1 42 POTENTIAL.
FT CHAIN 43 485 BETA-1,4-XYLANASE.
SQ SEQUENCE 485 AA; 51209 MW; 1E0BFD54C0B95A27 CRC64;

Query Match 45.0%; Score 244.5; DB 2; Length 485;
Best Local Similarity 49.5%; Pred. No. 2.3e-12;
Matches 46; Conservative 14; Mismatches 32; Indels 1; Gaps 1;

QY 3 GCRATVYVNSDMSGFTATVTNTGSRATSGWTVAWSFGGNOTVNYNWTALTQSGASV 62
Db 382 GCOVLGKV-NQNTGFTANVTNKNTSSAPVDGWTLTFSFGQOVTQANSSVTQSGSAV 440

QY 63 TATNLSYNNVIOPGQSTTFGNGSYSGTNTAPT 95
Db 441 TVRNAPWNGSIPAGCTAOFCPNGSHGTGNTAAPT 473

RESULT 8
Q8RQP6 PRELIMINARY; PRT; 623 AA.
AC Q8RQP6;
DT 01-JUN-2002 (TReMBLrel. 21, Created)
DT 01-JUN-2002 (TReMBLrel. 21, Last sequence update)
DE Chitinase III (EC 3.2.1.14) (Fragment).
GN CHI 111.
OS Streptomyces griseus.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1911;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=HUT 6037;
RA Watanabe T.;
RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=HUT 6037;
RA Kawase T., Kanai R., Ohno T., Tanabe T., Nikaidou N., Miyashita K.,
RA Mitutomi M., Watanabe T.;
RT "Identification of three family 18 chitinase genes of Streptomyces
RT griseus HUT6037.";
RL Chitin and chitosan research 7:241-251(2001).
DR EMBL; AB081807; BAB86377.1; -.
KW Hydrolase; Glycosidase.
FT NON_TER 1 623
FT CHAIN 623 623
SQ SEQUENCE 623 AA; 64893 MW; 02AAAB520A6C7E1C CRC64;
```

```
Query Match 43.3%; Score 235; DB 2; Length 623;
Best Local Similarity 50.6%; Pred. No. 1.8e-11;
Matches 44; Conservative 9; Mismatches 34; Indels 0; Gaps 0;

QY 6 ATYVNSDMSGFTATVTNTGSRATSGWTVAWSFGGNOTVNYNWTALTQSGASVTAT 65
Db 53 ATYTKSDMSGFEGQWTKNTCTTALSSWTIEWDPPSGTTAGSAMDATLTSSGHHYAK 112

QY 66 NLSYNNVIOPGQSTTFGNGSYSGTNT 92
Db 113 NLSWNGSVAPGASISFGNGTSGSPT 139

RESULT 9
Q54331 PRELIMINARY; PRT; 381 AA.
ID Q54331;
AC Q54331;
DT 01-NOV-1996 (TReMBLrel. 01, Created)
DT 01-NOV-1996 (TReMBLrel. 01, Last sequence update)
DE Cellulase B precursor.
DE CELB.
OS Streptomyces lividans.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1916;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=66;
RX MEDLINE=94288649; PubMed=8017952;
RA Witmann S., Sharek F., Kluepfel D., Morosoli R.;
RT "Purification and characterization of the CelB endoglucanase from
RT Streptomyces lividans 66 and DNA sequence of the encoding gene.";
RL Appl. Environ. Microbiol. 60:1701-1703(1994).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=66;
RA Sharek F.;
RL Submitted (OCT-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL; U04629; AAB71950.1; -.
DR HSSP; P07986; 1EXP.
DR InterPro; IPR001919; Bac_cellose-bind.
DR InterPro; IPR002594; GH_12.
DR Pfam; PF00553; CBM_2; 1.
DR Pfam; PF01670; Glyco_hydro_12; 1.
DR ProDom; PD004316; GH_12; 1.
DR PROSITE; PS00561; CBD_BACTERIAL; UNKNOWN_1.
KW Signal.
FT SIGNAL 1 40 POTENTIAL.
FT CHAIN 41 381 CELLULASE B.
SQ SEQUENCE 381 AA; 39239 MW; A7E99BF590FA24EC CRC64;

Query Match 42.1%; Score 228.5; DB 2; Length 381;
Best Local Similarity 48.9%; Pred. No. 3.6e-11;
Matches 45; Conservative 10; Mismatches 36; Indels 1; Gaps 1;

QY 4 CRATVYVNSDMSGFTATVTNTGSRATSGWTVAWSFGGNOTVNYNWTALTQSGASVT 63
Db 279 CAVSYGTNV-WQDGFADVTVTNTGAPVDGWOLATLPSGQRITNAWNASLTPSGSVT 337

QY 64 ATNLSYNNVIOPGQSTTFGNGSYSGTNTAPT 95
Db 338 ATGASHNARIAPGSLSPFGQCTYGCFAFAEPT 369

RESULT 10
Q9RJY3 PRELIMINARY; PRT; 381 AA.
ID Q9RJY3;
AC Q9RJY3;
DT 01-MAY-2000 (TReMBLrel. 13, Created)
DT 01-MAY-2000 (TReMBLrel. 13, Last sequence update)
DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)
```


RP SEQUENCE FROM N.A.
RA Vetrivel K.S., Pandian S.K., Chaudary U., Dharmalingam K.:
RT "Purification, cloning and DNA sequence analysis of a chitinase from
RT an overproducing mutant of Streptomyces peucetius defective in
RT daunorubicin biosynthesis."
RL Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: BELONGS TO CHITINASE CLASS II (FAMILY 18 OF GLYCOSYL
CC HYDROLASES).
CC -1- SIMILARITY: CONTAINS 1 FIBRONECTIN TYPE III-LIKE DOMAIN.
DR EMBL: AF206633; AAF43629.1; -.
DR HSSP: P07986; 1EXG.
DR InterPro: IPR001919; Bac_cellose-bind.
DR InterPro: IPR001579; Chitinase_18/2.
DR InterPro: IPR002048; EF-hand.
DR InterPro: IPR003961; FN_III.
DR InterPro: IPR001223; Glyco_hydro_18.
DR Pfam: PF00553; CBM_2; 1.
DR Pfam: PF00041; fn3; 1.
DR Pfam: PF00704; Glyco_hydro_18; 1.
DR ProDom: PD000471; Glyco_hydro_18; 1.
DR SMART: SM00060; FN3; 1.
DR PROSITE: PS00361; CBD_BACTERIAL; 1.
DR PROSITE: PS01095; CHITINASE_18; 1.
DR PROSITE: PS00018; EF_HAND; UNKNOWN_1.
KW Glycosidase; Hydrolase.
SQ SEQUENCE 629 AA; 65410 MW; 11E6214DFE8DC0F8 CRC64;

Query Match 41.3%; Score 224; DB 2; Length 629;
Best Local Similarity 43.9%; Pred. No. 1.4e-10;
Matches 43; Conservative 12; Mismatches 35; Indels 8; Gaps 1;

QY 1 GVGC-----RATVVVNSDMGSGFTATVTVTNGSRATSGMTVAWSFGGNGQTVTNYN 52
DB 44 GLACPAEAATSATATYTKTSDMGSGFGGWTVKNTGTTTINSWTVEWDFPFGKVTSAWD 103

QY 53 TALTOSGASVTATNLVNNVVGOSTTFEGNGSYCT 90
DB 104 ATVTNSGDHWTAKNVGNWGTLPAGTSVFGNGSGGS 141

RESULT 13
Q08166 PRELIMINARY; PRT; 974 AA.
AC Q08166
DT 01-NOV-1996 (TremBLrel. 01, Created)
DT 01-NOV-1996 (TremBLrel. 01, Last sequence update)
DT 01-JUN-2002 (TremBLrel. 21, Last annotation update)
DE Cellulase 1 precursor (EC 3.2.1.4) (Endoglucanase)
DE (Endo-1,4-beta-glucanase) (Avicelase).
GN Fl.
OS Thermomonospora fusca.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptosporangineae; Nocardiopsaceae; Thermobifida.
OX NCBI_TaxID=2021;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=YX;
RA Jung E.D., Lao G., Irwin D., Barr B., Benjamin A., Wilson D.B.;
RL Submitted (JUL-1993) to the EMBL/GenBank/DBJ databases.
CC -1- CATALYTIC ACTIVITY: ENDOHYDROLYSIS OF 1,4-BETA-D-GLUCOSIDIC
CC LINKAGES IN CELLULOSE.
CC -1- SIMILARITY: BELONGS TO CELLULOSE FAMILY E (FAMILY 9 OF GLYCOSYL
CC HYDROLASES).
CC EMBL: L20094; AAC06387.1; -.
DR HSSP: P07986; 1EXG.
DR InterPro: IPR001919; Bac_cellose-bind.
DR InterPro: IPR003305; CBM_CenC.
DR InterPro: IPR004197; celD_N.
DR InterPro: IPR003961; FN_III.
DR InterPro: IPR001701; GH_9.
DR InterPro: IPR000566; Lipocin_cytFABP.
DR InterPro: IPR001230; Prenyl_site.
DR Pfam: PF00553; CBM_2; 1.

DR Pfam: PF02018; CBM_4_9; 1.
DR Pfam: PF02927; celD_N; 1.
DR Pfam: PF00041; fn3; 1.
DR Pfam: PF00759; Glyco_hydro_9; 1.
DR SMART: SM00060; FN3; 1.
DR PROSITE: PS00698; GLYCOSYL_HYDROL_F9_2; 1.
DR PROSITE: PS00213; LIPOCALIN; UNKNOWN_1.
DR PROSITE: PS00294; PRENYLATION; UNKNOWN_1.
KW Cellulose degradation; Hydrolase; Glycosidase; Signal.
FT SIGNAL 1 32 POTENTIAL.
FT CHAIN 33 974 CELLULOSE 1.
FT DOMAIN 870 974 CELLULOSE-BINDING.
FT ACT_SITE 673 723 BY SIMILARITY.
FT ACT_SITE 723 732 BY SIMILARITY.
FT ACT_SITE 732 732 BY SIMILARITY.
SQ SEQUENCE 974 AA; 104577 MW; 17FE733040A83C CRC64;

Query Match 40.5%; Score 220; DB 2; Length 974;
Best Local Similarity 46.7%; Pred. No. 4.7e-10;
Matches 43; Conservative 15; Mismatches 32; Indels 2; Gaps 2;

QY 4 CRATVVNSDMGSGFTATVTVTNGSRATSGMTVAWSFGGNGQTVTNYNTALTOSGASVT 63
DB 871 CAVTYQTN-DWPGGFTASVTLTNTGTPWDSWELRFTFPSCQTVSHGWSANWQSGSDVT 929

QY 64 ATNLVNNVIOIP-GOSTTFEGNGSYCTNTP 94
DB 930 ATSLPWNQSVPPCGGSGVNI GFNGTWCNSNTP 961

RESULT 14
Q9S675 PRELIMINARY; PRT; 270 AA.
AC Q9S675
DT 01-MAY-2000 (TremBLrel. 13, Created)
DT 01-MAY-2000 (TremBLrel. 13, Last sequence update)
DT 01-JUN-2002 (TremBLrel. 21, Last annotation update)
DE Chitinase (Fragment).
GN CHI.
OS Streptomyces coelicolor.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1902;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=M145;
RA Kormanec J., Homerova D., Sevcikova B.;
RT "Two-component regulatory system in Streptomyces coelicolor A3(2).";
RL Submitted (OCT-1998) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: CONTAINS 1 FIBRONECTIN TYPE III-LIKE DOMAIN.
DR EMBL: AF102271; AAD16453.1; -.
DR HSSP: P07986; 1EXG.
DR InterPro: IPR001919; Bac_cellose-bind.
DR InterPro: IPR003961; FN_III.
DR InterPro: IPR001223; Glyco_hydro_18.
DR Pfam: PF00553; CBM_2; 1.
DR Pfam: PF00041; fn3; 1.
DR Pfam: PF00704; Glyco_hydro_18; 1.
DR ProDom: PD000471; Glyco_hydro_18; 1.
DR SMART: SM00060; FN3; 1.
DR PROSITE: PS00361; CBD_BACTERIAL; 1.
KW Hydrolase.
FT NON_TER 270 270
SQ SEQUENCE 270 AA; 27226 MW; D3361F55A258269A CRC64;

Query Match 39.2%; Score 213; DB 2; Length 270;
Best Local Similarity 45.9%; Pred. No. 4.6e-10;
Matches 39; Conservative 13; Mismatches 33; Indels 0; Gaps 0;

QY 6 ATVVVNSDMGSGFTATVTVTNGSRATSGMTVAWSFGGNGQTVTNYNTALTOSGASVTAT 65
DB 36 ATFAKTSMDGTGGGSGWTVKNTGTTSLSSWTVEWDFPTGKTVTSAMDATVTNSGDHWITAK 95

```

Oy 66 NLSYNNVIOPGSGTTFGFGNGSYSGT 90
I: I : I I : I I I I I I I :
Db 96 NVGWNGLTAPGASVSFGFGNGSGPGS 120

RESULT 15
Q929M8 PRELIMINARY; PRT; 609 AA.
ID Q929M8;
AC 01-MAY-1999 (TrEMBLrel. 10, Created)
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE CHIC (Chitinase C) (Fragment).
GN CHIC OR SC05376 OR 25C6G5.20C.
OS Streptomyces coelicolor.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1902;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RA Saito A., Fujii T., Yoneyama T., Redenbach M., Ohno T., Watanabe T.,
RA Miyashita K.;
RT "Structure of seven different chitinase genes from Streptomyces
RL coelicolor A3(2).";
RL Submitted (AUG-1998) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RA Seeger K.J., Harris D.;
RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RA Bentley S.D., Parkhill J., Barrell B.G., Rajandream M.A.;
RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RX MEDLINE=97000351; PubMed=8843436;
RA Redenbach M., Kieser H.M., Denapante D., Eichner A., Cullum J.,
RA Kinashi H., Hopwood D.A.;
RT "A set of ordered cosmids and a detailed genetic and physical map for
RL the 8 Mb Streptomyces coelicolor A3(2) chromosome.";
RL Mol. Microbiol. 21:77-96(1996).
RN [5]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2) / M145;
RA Bentley S.D., Chater K.F., Cerdeno-Tarraga A.-M., Challis G.L.,
RA Thomson N.R., James K.D., Harris D.E., Ouali M.A., Kieser H.,
RA Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,
RA Cronin A., Fraser A., Goble A., Hidalgo J., Hornsby T., Howarth S.,
RA Huang C.-H., Kieser T., Larke L., Murphy L., Oliver K., O'Neil S.,
RA Rabinowitsch E., Rajandream M.A., Rutherford K., Rutter S.,
RA Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K.,
RA Warren T., Wietzorrek A., Woodward J., Barrell B.G., Parkhill J.,
RA Hopwood D.A.;
RT "Complete genome sequence of the model actinomycete Streptomyces
RL coelicolor A3(2).";
RL Nature 417:141-147(2002).
CC -1- SIMILARITY: BELONGS TO CHITINASE CLASS II (FAMILY 18 OF GLYCOSYL
CC HYDROLASES).
CC -1- SIMILARITY: CONTAINS 1 FIBRONECTIN TYPE III-LIKE DOMAIN.
DR EMBL; AB017010; BAA75644.1; -.
DR EMBL; AL359152; CAB94547.1; -.
DR HSSP; P07986; IEXG
DR InterPro; IPR001919; Bac_celose-bind.
DR InterPro; IPR001579; Chitinase_18/2.
DR InterPro; IPR002048; EF-hand.
DR InterPro; IPR003961; FN_III.
DR InterPro; IPR001223; Glyco_hydro_18.
DR Pfam; PF00553; CBM_2; 1.
DR Pfam; PF00041; fn3; 1.

```

```

DR Pfam; PF00704; Glyco_hydro_18; 1.
DR ProDom; PD000471; Glyco_hydro_18; 1.
DR SMART; SM00060; FN3; 1.
DR PROSITE; PS00561; CBD_BACTERIAL; 1.
DR PROSITE; PS01095; CHITINASE_18; 1.
DR PROSITE; PS00018; EF_HAND; UNKNOWN_1.
KW Glycosidase; Hydrolase.
FT NON_TER 1
FT NON_TER 609
SQ SEQUENCE 609 AA; 63621 MW; 576F69F1A4343570 CRC64;

Query Match 39.2%; Score 213; DB 16; Length 609;
Best Local Similarity 45.9%; Pred. No. 1.1e-09;
Matches 39; Conservative 13; Mismatches 33; Indels 0; Gaps 0;

Oy 6 ATYVVVNSDWGSGFTATVTVTNGSRATSWTVAWSFGGNGHQTVTNYNNTALTOSGASVYAT 65
I: I : I I : I I I I : I I I I I I I : I I : I I I I
Db 36 ATFAKTSDMGTGFGSGWTVKNTGTTSLSSWTVEWDFPTGCTKVTSAWDATVTNSGDIIWTAK 45

Oy 66 NLSYNNVIOPGSGTTFGFGNGSYSGT 90
I: I : I : I I : I I I I I I I :
Db 96 NVGWNGLTAPGASVSFGFGNGSGPGS 120

Search completed: November 13, 2002, 11:54:12
Job time : 10.3599 secs

```


PR 26-SEP-1989; 89US-0412434.
 PR 27-JAN-1992; 92US-0826089.
 PR 21-SEP-1993; 93US-0125115.
 XX
 PA (MIDE) MIDWEST RES INST.
 XX
 XX Himmel ME, Laymon RA, Thomas SR;
 XX
 XX WPI; 1996-105843/11.
 DR N-PSDB; AAT12337.
 DR
 XX New isolated DNA encoding endoglucanase - obtd from Acidothermus
 PT cellulolyticus, used for prodn of the enzyme for use in cellulose
 PT hydrolysis.
 XX
 XX Claim 1: Page 22; 34pp; English.
 PS
 XX Acidothermus cellulolyticus E1 endoglucanase (AAR89927) is useful
 CC for hydrolysing cellulosic biomass to sugars for simultaneous or
 CC subsequent fermentation to ethanol. It shows optimal activity at
 CC 83 deg C. The amino acid sequence of the mature enzyme was deduced
 CC from an E1 endoglucanase gene (AAT12337) cloned from A. cellulolyticus.
 CC Putative signal peptides were also identified (see also AAR89928-29).
 CC Cloning of this gene allows large-scale, low-cost prodn. of
 CC recombinant E1 endoglucanase, using pref. Saccharomyces, Zymomonas
 CC or E.coli hosts.
 CC
 XX
 XX Sequence 521 AA;
 SQ
 Query Match 73.5%; Score 399; DB 17; Length 521;
 Best Local Similarity 69.3%; Pred. No. 9.4e-30;
 Matches 70; Conservative 13; Mismatches 18; Indels 0; Gaps 0;
 QY 1 GVCGRATYVNSDMGSGFTATVTNTGSRATSGWTVAWSFGGNOTVTNYWNTALTOSGA 60
 DB 421 GACTASYQVNSDMGNGFTVTVAVTNSGVSATKTWTVSGTGTITNSWNAAVTONGQ 480
 QY 61 SVTATNLISNNYVIOPGOSTTFFGNGSYSGTNTAPTLTCTAS 101
 DB 481 SVTARNNSNNYVIOPGQNTTFFGFOASYTGSNAAPTACAAAS 521
 RESULT 2
 AAB48786
 ID AAB48786 standard; protein: 521 AA.
 XX
 AC AAB48786;
 XX
 DT 09-MAR-2001 (first entry)
 XX
 DE Acidothermus cellulolyticus E1 endoglucanase mutant, W42R.
 XX
 KW E1 endoglucanase; glycosyl hydrolase; soluble substrate;
 KW cellulose hydrolysis; ethanol production; fermentation;
 KW mutant; mutein.
 XX
 OS Acidothermus cellulolyticus.
 OS Synthetic.
 XX
 PN WO200070031-A1.
 XX
 PD 23-NOV-2000.
 XX
 PF 19-MAY-2000; 2000WO-US13971.
 XX
 PR 19-MAY-1999; 99US-0134925.
 XX
 XX (MIDE) MIDWEST RES INST.
 PA
 XX Himmel ME, Adney WS, Baker JO, Vinzant TB, Thomas SR, Sakon J;
 PI Decker SR;
 XX
 XX WPI; 2001-061226/07.
 DR

XX
 PT Preparation of glycosyl hydrolase with an increased catalytic activity
 PT on insoluble substrate -
 XX
 PS Claim 16; Page 22-24; 30pp; English.
 XX
 XX The invention relates to a method for making glycosyl hydrolase mutants
 CC with increased catalytic activity with either insoluble or soluble
 CC cellulose substrates relative to the wild-type enzyme. The method for
 CC making a glycosyl hydrolase with increased insoluble substrate catalytic
 CC activity comprises replacing an active site-associated amino glycosyl-
 CC stabilising amino acid with an amino acid that does not strongly bind a
 CC disaccharide product in the active site. Conversely, the method for
 CC making a glycosyl hydrolase with increased soluble substrate catalytic
 CC activity comprises replacing a hydrophobic substrate-binding amino acid
 CC with a positively charged residue. The invention also discloses mutants
 CC of Acidothermus cellulolyticus E1 endoglucanase (AAB48786-B48788)
 CC produced according to the method of the invention. The Y245C mutant
 CC (AAB48786) has improved activity with insoluble substrates, and the W42R
 CC (AAB48786) and Y82R (AAB48787) mutants have improved activity with
 CC soluble substrates. The invention also encompasses DNA encoding these
 CC mutants. The glycosyl hydrolases of the invention are used as catalysts
 CC for cellulose hydrolysis to produce sugars that can be fermented to
 CC produce fuels such as ethanol. The present sequence represents the
 CC Acidothermus cellulolyticus E1 endoglucanase W42R mutant.
 XX
 XX Sequence 521 AA;
 SQ
 Query Match 73.5%; Score 399; DB 22; Length 521;
 Best Local Similarity 69.3%; Pred. No. 9.4e-30;
 Matches 70; Conservative 13; Mismatches 18; Indels 0; Gaps 0;
 QY 1 GVCGRATYVNSDMGSGFTATVTNTGSRATSGWTVAWSFGGNOTVTNYWNTALTOSGA 60
 DB 421 GACTASYQVNSDMGNGFTVTVAVTNSGVSATKTWTVSGTGTITNSWNAAVTONGQ 480
 QY 61 SVTATNLISNNYVIOPGOSTTFFGNGSYSGTNTAPTLTCTAS 101
 DB 481 SVTARNNSNNYVIOPGQNTTFFGFOASYTGSNAAPTACAAAS 521
 RESULT 3
 AAB48787
 ID AAB48787 standard; protein: 521 AA.
 XX
 AC AAB48787;
 XX
 DT 09-MAR-2001 (first entry)
 XX
 DE Acidothermus cellulolyticus E1 endoglucanase mutant, Y82R.
 XX
 KW E1 endoglucanase; glycosyl hydrolase; soluble substrate;
 KW cellulose hydrolysis; ethanol production; fermentation;
 KW mutant; mutein.
 XX
 OS Acidothermus cellulolyticus.
 OS Synthetic.
 XX
 PN WO200070031-A1.
 XX
 PD 23-NOV-2000.
 XX
 PF 19-MAY-2000; 2000WO-US13971.
 XX
 PR 19-MAY-1999; 99US-0134925.
 XX
 XX (MIDE) MIDWEST RES INST.
 PA
 XX Himmel ME, Adney WS, Baker JO, Vinzant TB, Thomas SR, Sakon J;
 PI Decker SR;
 XX
 XX WPI; 2001-061226/07.
 DR

XX PA (MIDE) MIDWEST RES INST.
XX PI Laymon RA, Adney WS, Thomas SR, Himmel ME;
XX PR WPI: 2000-087663/08.
DR N-PSDB; AA25924.
XX Isolated domains of Acidothermus cellulolyticus E1 endoglucanase useful
PT for labeling or modifying a cellulose and for purifying or immobilizing
PT a binding domain fusion protein to cellulose .
XX PS Claim 1: Fig 2: 85pp; English.
XX This sequence represents the Acidothermus cellulolyticus E1
CC endoglucanase, which is a beta-1,4-endoglucanase, or endocellulase. The
CC cellulose binding domain (CBD) of E1 endoglucanase, and nucleotides which
CC encode it are specifically claimed. The CBD is believed to be roughly
CC wedge-shaped; as the CBD binds to cellulose, it is thought that the wedge
CC tip is inserted between the microfibrils of the cellulose fibre,
CC disrupting the crystalline structure, and making the cellulose linkages
CC more accessible to the catalytic domain of the E1 endoglucanase. The E1
CC endoglucanase CBD is useful in labelling or modifying the surface of
CC cellulose or other polysaccharides. Such modified cellulose can then be
CC used in textile, pulp, paper, chemical and pharmaceutical industries.
CC CBDs can be used in affinity purification of CBD-fusion proteins, and can
CC also be used to immobilise the CBD-fusion proteins to a cellulose
CC support. CBD-fusion proteins can be used to modify the chemical or
CC physical properties of a cellulose or polysaccharide matrix column and to
CC modify (e.g., roughen or disrupt) a cellulose or polysaccharide fibre.
CC The CBD of the E1 endoglucanase exhibits greater stability at pH 4-8 and
CC has an optimum temperature for stability of 83 degrees Celsius which is
CC not found in CBDs from non-thermophilic organisms.
XX SQ Sequence 562 AA;
Query Match 73.5%; Score 399; DB 21; Length 562;
Best Local Similarity 69.3%; Pred. No. 1e-29;
Matches 70; Conservative 13; Mismatches 18; Indels 0; Gaps 0;
QY 1 GVGCRTATVYVNSDVGSGFTATVTNTGSRATSGTVAWSFGGNOTVTNYNTALTQSGA 60
DB 462 GARTASYQVNSDVGNGFTVTVAVNTSGSVATKTWTVSWTFGGNOTITNSNNAAYTONGQ 521
QY 61 SVTATNLNYYNNVIOPGOSTTFFGNGSYSGTNTAPTTLCTAS 101
DB 522 SVTARNMSYNNVIOPGONTTFFGQASVTGSGNAAPTACAAS 562
RESULT 6
AAU79549
XX AAU79549 standard; Protein; 562 AA.
AC AAU79549;
XX 24-SEP-2002 (first entry)
DE A. cellulolyticus cellulase EI beta-1,4-endoglucanase precursor.
XX Gene; ds; transgenic; plant; lignocellulose; cellulase; ligninase;
KW fermentable sugar; ethanol; fermentation; silage; feed; fuel;
KW Industrial chemical; biodegradation; chloroaromatic;
KW environmental pollutant; EI beta-1,4-endoglucanase; el.
XX Acidothermus cellulolyticus.
OS
XX Key Location/Qualifiers
FH Peptide 1..41
FT /label= Leader_sequence
FT 42..562
FT /label= Mature_cellulase
XX W0200234926-A2.

XX PD 02-MAY-2002.
XX PF 18-OCT-2001; 2001WO-US32538.
XX PR 20-OCT-2000; 2000US-242408P.
XX PA (UNMS) UNIV MICHIGAN STATE.
XX PI Sticklen MB, Dale BE, Magbool S;
XX WPI: 2002-489947/52.
DR N-PSDB; ABK86729.
XX Producing transgenic plants which after harvest degrade lignin and
PT cellulose to fermentable sugars, by mating transgenic plant comprising
PT DNA encoding cellulase with transgenic plant comprising DNA encoding
PT ligninase .
XX Disclosure; Page 96-99; 126pp; English.
XX The invention discloses the production of a transgenic plant which
CC degrades lignocellulose when the plant is grown. It comprises the
CC production of the transgenic plant including cellulase and ligninase by
CC mating a transgenic plant, containing a DNA encoding a cellulase, and a
CC transgenic plant, containing a DNA encoding a ligninase, where both genes
CC are operably linked to a nucleotide sequence encoding a signal peptide
CC which targets the fusion protein to an organelle of the plant,
CC particularly chloroplasts. The method is useful for producing a
CC transgenic plant (e.g. maize) which degrades lignocellulose when the
CC plant is grown to produce a plant material. This material is useful for
CC converting lignocellulose, in a plant material, to fermentable sugars
CC which are then fermented to ethanol. The transgenic plants also provide a
CC plentiful and inexpensive source of fungal or bacterial cellulases and
CC ligninases which can be used in the production of ethanol. They can also
CC be used for pre-treating silage to increase the energy value of
CC lignocellulosic feeds for cows and other ruminant animals, pre-treating
CC lignocellulosic biomass for fermentative conversion to fuels and
CC industrial chemicals, and biodegradation of chloroaromatic environmental
CC pollutants. The protein sequence presented is the A. cellulolyticus
CC cellulase EI beta-1,4-endoglucanase precursor.
XX SQ Sequence 562 AA;
Query Match 73.5%; Score 399; DB 23; Length 562;
Best Local Similarity 69.3%; Pred. No. 1e-29;
Matches 70; Conservative 13; Mismatches 18; Indels 0; Gaps 0;
QY 1 GVGCRTATVYVNSDVGSGFTATVTNTGSRATSGTVAWSFGGNOTVTNYNTALTQSGA 60
DB 462 GARTASYQVNSDVGNGFTVTVAVNTSGSVATKTWTVSWTFGGNOTITNSNNAAYTONGQ 521
QY 61 SVTATNLNYYNNVIOPGOSTTFFGNGSYSGTNTAPTTLCTAS 101
DB 522 SVTARNMSYNNVIOPGONTTFFGQASVTGSGNAAPTACAAS 562
RESULT 7
AAR90715
ID AAR90715 standard; Protein; 596 AA.
XX AAR90715;
AC AAR90715;
XX 06-APR-1996 (first entry)
DT Thermostable cellulase-E3.
XX Thermostable cellulase-E3.
XX Cellulase-E3; Thermomonospora fusca; thermostable; plasmid pS24;
KW plasmid pS26; Escherichia coli; cloning; Streptomyces lividans;
KW papain; catalytic domain; cellulose; chitosan; protease resistance;
KW synergism; cellobiohydrolase; beta-glucosidase; saccharification;
KW surfactant additive; paper recycling; delinking; paper refining.
XX


```
FT FT /label= eglin_C
FT 72..79
FT /label= linker
FT 80..212
FT /label= cellulose_binding_domain
FT /note= "from Cellulomonas fimi Exg protein "
FT 72..73
FT /label= KEX2/KEX1_recognition
FT EP467839-A.
PN
XX
XX 22-JAN-1992.
XX
XX 09-JUL-1991; 91EP-0810543.
XX
XX 18-JUL-1990; 90GB-0015825.
XX
XX (CIBA ) CIBA GEIGY AG.
XX
XX Helm J, Seeboth P, Takabayashi K:
XX WPI 1992-026641/04.
XX
XX In vitro processing of fusion protein - using yeast endo:protease
PT ysc F and carboxy:peptidase ysc alpha, for prodn. of biologically
PT active mature protein
XX
XX Disclosure; Page 30; 35pp; English.
XX
XX The sequence (SEQ ID NO: 4) is encoded by an insert from
CC PDP34GAPDH- eglinex-1, a vector constructed for the expression of
CC the fusion protein in Saccharomyces cerevisiae. The protein has a
CC formula P-L-T, in which P is the eglin C polypeptide (Rink H.
CC et al. 1984 Nucl. Acids Res. 12 6369-6387), L a linker incorporating
CC the KEX2/KEX1 recognition site, and T the cellulose binding domain
CC of the C. fimi Exg protein (including an extra proline) encoded by
CC the cex gene (Neill G.O. et al. 1986 Gene 44 325-330). On expression
CC the crude fusion protein is purified by cellulose affinity
CC chromatography. Fractions contg. the protein are then treated with
CC sol. yscF resulting in the cleavage prod. eglin C-Lys-Arg which is
CC then digested with sol. ysc alpha to remove the Lys-Arg tail.
CC The obtd. eglin C is identical in structure to natural eglin C
CC as evidenced by HPLC.
CC See also AAR20469-71.
XX
XX SQ Sequence 212 AA:
Query Match 45.0%; Score 244.5; DB 13; Length 212;
Best Local Similarity 49.5%; Pred. No. 1.5e-15;
Matches 46; Conservative 14; Mismatches 32; Indels 1; Gaps 1;
QY 3 GCRATYVNSDMGSGFTATVTNTGSRATSGMTVAWSFGNGNOTVTNWNLTALTSQGSV 62
DB 109 GCQVLGVN-QWNTGFTANVTYKNTSAPVDGWTLTFFSPSGQOVTOAWSSTVTSQSAV 167
QY 63 TATNLSYNNVIOPCQSTTFGNGSYSGTNAPT 95
DB 168 TVRNAPMNGSIPAGGTAQFGNGSHTGTNAAPT 200
RESULT 10
AAB31372
ID AAB31372 standard; Protein; 467 AA.
XX
XX AAB31372;
XX
XX 20-APR-2001 (first entry)
XX
XX Amino acid sequence of protein L/CBD cex/ER retaining peptide fusion.
XX
XX Protein production; food processing; protein antibiotic; feed enzyme;
KW protein L: CBD cex protein; cell signal peptide.
XX
```

```
OS Synthetic.
XX
XX PN WO20007174-A1.
XX
XX PD 21-DEC-2000.
XX
XX PF 07-JUN-2000; 2000WO-IL00330.
XX
XX PR 10-JUN-1999; 99US-0329234.
XX
XX (CBDT-) CBD TECHNOLOGIES LTD.
PA (YISS ) YISSUM RES DEV CO HEBREW UNIV JERUSALEM.
XX
XX Shani Z, Shoseyov O;
XX
XX WPI; 2001-112219/12.
DR N-PSDB; AAF24730.
XX
XX Expressing and isolating recombinant protein in a plant, useful for
PT producing large quantities of recombinant proteins, by expressing a
PT fusion protein including a cellulose binding peptide fused to a
PT recombinant protein -
XX
XX Example; Fig 2a; 87pp; English.
XX
XX The specification describes a method for expressing and isolating
CC a recombinant protein in a plant. The method comprising expressing a
CC fusion protein including the recombinant protein and a cellulose
CC binding peptide fused to it, where the fusion protein is
CC compartmentalised and sequestered within plant cells, plant derived
CC tissue or cultured plant cells. The method is useful for obtaining large
CC quantities of the recombinant proteins and protein products in a simple
CC and cost-effective manner. Recombinant proteins may be used commercially,
CC such as in the food processing industry, e.g. glucoamylases and glucose
CC isomerases are used for converting starch to high fructose corn syrup,
CC manufactures for the hydrolysis of high molecular weight proteins and in
CC pectin hydrolysis in food industry, lipases for cleaving ester linkage
CC in triglycerides, and for effluent treatment. The recombinant proteins
CC may further be used to produce protein antibiotics, which can be used
CC in healing processes, and to produce animal feed enzymes. The present
CC sequence represents a fusion protein of the invention, and comprises a
CC fusion of a cell signal peptide, protein L, CBD cex, and an endoplasmic
CC reticulum retaining peptide.
XX
XX SQ Sequence 467 AA:
Query Match 44.7%; Score 242.5; DB 22; Length 467;
Best Local Similarity 49.5%; Pred. No. 6.2e-15;
Matches 46; Conservative 14; Mismatches 32; Indels 1; Gaps 1;
QY 3 GCRATYVNSDMGSGFTATVTNTGSRATSGMTVAWSFGNGNOTVTNWNLTALTSQGSV 62
DB 360 GCQVLGVN-QWNTGFTAQVTVKNTSAPVDGWTLTFFSPSGQOVTOAWSSTVTSQSAV 418
QY 63 TATNLSYNNVIOPCQSTTFGNGSYSGTNAPT 95
DB 419 TVRNAPMNGNIPAGGTAQFGNGSHTGTNAAPT 451
RESULT 11
AAB31373
ID AAB31373 standard; Protein; 509 AA.
XX
XX AAB31373;
XX
XX 20-APR-2001 (first entry)
XX
XX Amino acid sequence of Sig/CBD cex/ER retaining peptide fusion.
XX
XX Protein production; food processing; protein antibiotic; feed enzyme;
KW protein L; CBD cex protein; cell signal peptide.
XX
```

OS Synthetic.
 PN WO200077174-A1.
 XX 21-DEC-2000.
 XX 07-JUN-2000; 2000WO-IL00330.
 XX 10-JUN-1999; 99US-0329234.
 XX (CBMT-) CBD TECHNOLOGIES LTD.
 PA (VISS) YISSUM RES DEV CO HEBREW UNIV JERUSALEM.
 XX Shani Z, Shoseyov O;
 XX WPI; 2001-112219/12.
 DR N-PSDB; AAF24735.
 XX Expressing and isolating recombinant protein in a plant, useful for
 PT producing large quantities of recombinant proteins, by expressing a
 PT fusion protein including a cellulose binding peptide fused to a
 PT recombinant protein -
 XX Example: Fig 3a; 87pp; English.
 PS The specification describes a method for expressing and isolating
 CC a recombinant protein in a plant. The method comprising expressing a
 CC fusion protein including the recombinant protein and a cellulose
 CC binding peptide fused to it, where the fusion protein is
 CC compartmentalised and sequestered within plant cells, plant derived
 CC tissue or cultured plant cells. The method is useful for obtaining large
 CC quantities of the recombinant proteins and protein products in a simple
 CC and cost-effective manner. Recombinant proteins may be used commercially,
 CC such as in the food processing industry, e.g. glucoamylases and glucose
 CC isomerases are used for converting starch to high fructose corn syrup.
 CC proteinases for the hydrolysis of high molecular weight proteins and in
 CC manufacturing leather or alcoholic beverages, pectinesterases for
 CC pectin hydrolysis in food industry, lipases for cleaving ester linkage
 CC in triglycerides, and for effluent treatment. The recombinant proteins
 CC may further be used to produce protein antibiotics, which can be used
 CC in healing processes, and to produce animal feed enzymes. The present
 CC sequence represents a fusion protein of the invention, and comprises a
 CC fusion of a cell signal peptide, Sig (not specified), CBD cex, and an
 CC endoplasmic reticulum retaining peptide.
 XX
 XX Sequence 509 AA;
 S0 Query Match 44.7%; Score 242.5; DB 22; Length 509;
 Best Local Similarity 49.5%; Pred. No. 6.0e-15;
 Matches 46; Conservative 14; Mismatches 32; Indels 1; Gaps 1;
 QY 3 GCRTATVYVNSDWGSGFTATVTVTGTGSRATSGWTVAVSFGCNGQTVTVYNTALTQSGASV 62
 DB 41 GCQVLWGVN-QWNTGFTAQVTVKTKGAPVDGWTLTFTSPSGQQVTOAWSVTVTQSGSAV 99
 QY 63 TATNLVSNVNIQPGOSTTFGNGSYSGTNTAPT 95
 DB 100 TVRNAPWNGNIPAGGTAQFGQSGHTGTNAAPT 132
 RESULT 12
 ABB06928
 ID ABB06928 standard; Protein; 518 AA.
 XX ABB06928;
 AC ABB06928;
 XX 18-JUN-2002 (first entry)
 DT Streptomyces lividans EGI11-like cellulase.
 DE Cellulase; endoglucanase; EGI11; textile; feed additive; baking;
 XX food processing; grain wet milling; pulp; paper.
 XX Streptomyces lividans.
 OS WO9931255-A2.
 PN 24-JUN-1999.
 XX 14-DEC-1998; 98WO-US26552.
 XX 16-DEC-1997; 97US-0991720.

OS Micromonospora carbonacea.
 XX WO200155180-A2.
 PN 02-AUG-2001.
 XX 29-JAN-2001; 2001WO-CA00128.
 XX 27-JAN-2000; 2000US-0177711.
 XX (ECOP-) ECOPIA BIOSCIENCES INC.
 PA (FARN/) FARNET C.
 XX Staffa A, Zazopoulos E, Mercure S, Nowacki P;
 XX WPI; 2001-476185/51.
 DR N-PSDB; ABL50562.
 XX Novel isolated gene cluster encoding polypeptides involved in
 PT everninomicin biosynthesis useful for construction of everninomicin
 PT overproducing strains, and to allow chemical modifications of
 PT everninomicin to enhance certain properties -
 XX Claim 15; Fig 1; 181pp; English.
 PS ABL50555 to ABL50562 represent contigs 1 to 8 from the Micromonospora
 CC carbonacea everninomicin biosynthetic locus gene cluster. The contigs
 CC encode the protein sequences designated ORF (open reading frame) 1 to
 CC 49, given in ABB06881 to ABB06930. The gene cluster is useful for the
 CC construction of the everninomicin antibiotic in overproducing strains,
 CC and to allow chemical modifications of everninomicin to enhance certain
 CC properties via genetic manipulation or combinatorial biosynthesis. The
 CC gene cluster can be used to produce genetic systems and genes encoding
 CC novel enzyme activities, and avoid the problems of low yield and quality
 CC of everninomicins produced by chemical synthesis.
 XX
 XX Sequence 518 AA;
 S0 Query Match 42.5%; Score 231; DB 22; Length 518;
 Best Local Similarity 45.3%; Pred. No. 8.6e-14;
 Matches 48; Conservative 13; Mismatches 37; Indels 8; Gaps 3;
 QY 3 GCRTATVYVNSDWGSGFTATVTVTGTGSRATSGWTVAVSFGCNGQTVTVYNTALTQSGAS 61
 DB 34 GCSVAYTVQSQWGTGFCGNVAITNLGS-ALGWTLTFTDFPTSGQQVTCGWSATWSQSGTS 92
 QY 62 VTATNLVSNVNIQPGOSTTFGNGSYSGTNTAPT-----LTCTAS 101
 DB 93 VSAASLSWNGSLGTGGTGTITGFGWSGSGSNPVPKSFALNGTTCTGGS 138
 RESULT 13
 AAY06368
 ID AAY06368 standard; Protein; 381 AA.
 XX AAY06368;
 AC AAY06368;
 XX 06-SEP-1999 (first entry)
 DT Streptomyces lividans EGI11-like cellulase.
 DE Cellulase; endoglucanase; EGI11; textile; feed additive; baking;
 KW food processing; grain wet milling; pulp; paper.
 XX Streptomyces lividans.
 OS WO9931255-A2.
 PN 24-JUN-1999.
 XX 14-DEC-1998; 98WO-US26552.
 XX 16-DEC-1997; 97US-0991720.

XX (GEMV) GENENCOR INT INC.
 XX PA Bower BS, Fowler T, Phillips JI;
 XX PI WPI; 1999-395187/33.
 XX DR EGIII like cellulase
 XX PT Example; Fig 6; 47pp; English.
 XX PS The present polypeptide represents a full-length sequence of a
 CC novel EGIII-like cellulase of Streptomyces lividans (CeIB). It was
 CC deduced from a gene sequence isolated from genomic DNA using PCR
 CC primers (see AAX59180-91) based on conserved motifs (see AAY06325-29)
 CC of Trichoderma reesei EGIII cellulase and related enzymes. PCR
 CC has been used to identify novel EGIII-like enzymes, including the
 CC present protein, from bacterial and fungal sources (see AAY06331-70).
 CC The sequence shows homology to T. reesei EGIII (see AAY06330). Also
 CC provided by the invention are vectors, host cells and methods
 CC for the recombinant production of such enzymes, which can be used
 CC in the treatment of cellulose-containing textiles, as feed
 CC additives, in the treatment of wood pulp, in the reduction of
 CC biomass to glucose, in the stone washing of indigo dyed denim, or
 CC as laundry detergent components (all claimed).
 XX CC
 XX SQ Sequence 381 AA;
 Query Match 42.1%; Score 228.5; DB 20; Length 381;
 Best Local Similarity 48.9%; Pred. No. 1e-13;
 Matches 45; Conservative 10; Mismatches 36; Indels 1; Gaps 1;
 QY 4 CRATYVNSDMGSGFTATVTVTGSRATSGWTVAMSGGNGQTVNTYNTALTOSGASVT 63
 DB 279 CAVSYGTNV-WQDGFADVTVTGTPAPVDCWQLAFTLPSCQRITNAWNASLTPSSGSVT 337
 QY 64 ATNLSYNNVIQPGOSTTGFNGSYSGTNTAPT 95
 DB 338 ATGASHNARIAPGSLSGFGQTYGGAFAPET 369
 RESULT 14
 AABI4881
 ID AABI4881 standard; Protein: 381 AA.
 XX AC AABI4881;
 XX DT 21-NOV-2000 (first entry)
 XX DE Rhodothermus marinus EGIII-like cellulase.
 XX KW Rhodothermus marinus; Trichoderma reesei; endoglucanase III; EGIII;
 KW cellulase; mutant; enzyme stability; textile treatment;
 KW wood pulp treatment; feed additive; detergent.
 XX OS Rhodothermus marinus.
 XX PN WQ200037614-A2.
 XX PD -J-JUN-2000.
 XX PF 14-NOV-1999; 99WO-US26704.
 XX PR 16-NOV-1998; 98US-0216295.
 XX PS (GEMV) GENENCOR INT INC.
 XX PI Mitchinson C, Wendt DJ;
 XX DR WPI; 2000-482483/42.
 XX PT Novel endoglucanase III or endoglucanase III-like cellulase useful for
 PT treating textiles and wood pulp comprises a substitution or deletion at

PT specified positions in the wild form of endoglucanase III -
 XX Example 1; Fig 3; 52pp; English.
 XX CC The present sequence is a cellulase related to endoglucanase III (EGIII)
 CC from Trichoderma reesei. EGIII-like genes were isolated from genomic DNA
 CC libraries constructed from various microorganisms by PCR. The isolated
 CC genes showed significant homology to EGIII from T. reesei. Certain
 CC substitution and deletion mutations have been incorporated into EGIII and
 CC EGIII-like cellulases to produce variant enzymes with improved stability,
 CC e.g. increased resistance to temperature stress. The mutants may be used
 CC in textile and wood pulp treatment, as a feed additive, and for reducing
 CC biomass to glucose. They are also useful for stonewashing or indigo dyed
 CC denim and as an agent in laundry and dish detergents.
 XX CC
 XX SQ Sequence 381 AA;
 Query Match 42.1%; Score 228.5; DB 21; Length 381;
 Best Local Similarity 48.9%; Pred. No. 1e-13;
 Matches 45; Conservative 10; Mismatches 36; Indels 1; Gaps 1;
 QY 4 CRATYVNSDMGSGFTATVTVTGSRATSGWTVAMSGGNGQTVNTYNTALTOSGASVT 63
 DB 279 CAVSYGTNV-WQDGFADVTVTGTPAPVDCWQLAFTLPSCQRITNAWNASLTPSSGSVT 337
 QY 64 ATNLSYNNVIQPGOSTTGFNGSYSGTNTAPT 95
 DB 338 ATGASHNARIAPGSLSGFGQTYGGAFAPET 369
 RESULT 15
 AAU77433
 ID AAU77433 standard; Protein: 381 AA.
 XX AC AAU77433;
 XX DT 05-JUN-2002 (first entry)
 XX DE Streptomyces lividans CeIB EGIII-like cellulase.
 XX KW Endoglucanase III-like cellulase; EGIII-like;
 KW cellulose containing textile; enzyme.
 XX OS Streptomyces lividans CeIB.
 XX PN WQ200212464-A2.
 XX PD 14-FEB-2002.
 XX PF 31-JUL-2001; 2001WO-US23989.
 XX PR 04-AUG-2000; 2000US-0632426.
 XX PS (GEMV) GENENCOR INT INC.
 XX PI Mitchinson C, Ropp TH, Swanson BA;
 XX DR WPI; 2002-241750/29.
 XX PT Novel endoglucanase III (EGIII)-like cellulase variant comprising
 PT substitution/deletion at positions corresponding to specific residues
 PT in EGIII from Trichoderma reesei, useful for treating cellulose
 PT containing textile -
 XX Example 1; Fig 3; 41pp; English.
 XX CC The present invention relates to novel endoglucanase III (EGIII)-like
 CC cellulase variants which comprise a substitution or deletion at a
 CC position corresponding to one or more of residues W79, M154 and/or
 CC M118 in mature EGIII from the fungus, Trichoderma reesei. The variants
 CC are useful in the treatment of a cellulose containing textile. By
 CC substituting other amino acids for the native methionines at
 CC positions 79, 118 and 154 (sites where oxidation of the enzyme takes

CC place) in EGIII from T. reesei, oxidatively more stable enzymes are
CC obtained. The present sequence represents Streptomyces lividans CelB
CC EGIII-like cellulase.
XX

SO Sequence 381 AA:

Query Match 42.1%; Score 228.5; DB 23; Length 381;
Best Local Similarity 48.9%; Pred. No. 1e-13;
Matches 45; Conservative 10; Mismatches 36; Indels 1; Gaps 1;

OY 4 CRATVVNSDNGSGFTATVTNTGSRATSGTVAWSFGGNOTVTNYWNTALTQSGASVT 63

Db 279 CAVSYGTNV-WQDGTADVTNTGTAPVDCWQLAFTLPSCORITNAWNASLTPSSGSVT 337

OY 64 ATNLSYNNVIOPGQSTTFENGSGSYSGTNTAPT 95

Db 338 ATGASHNARIAPGSLSPFGQGYCGAFAPT 369

Search completed: November 13, 2002, 11:51:41
Job time : 10.0697 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 13, 2002, 11:49:46 ; Search time 3.1197 Seconds
(without alignments)
952.482 Million cell updates/sec

Title: US-09-917-378-5

Perfect score: 543

Sequence: 1 GVCCRTATVYVNDWGSFGFTA.....GFNGSYSGTNTAPLTCTAS 101

Scoring table: BLOSUM62

Gapop 10.0 , Capext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued_Patents_AR.*

- 1: /cgn2_6/ptodata/1/iaa/5A_COMB.pep.*
- 2: /cgn2_6/ptodata/1/iaa/5B_COMB.pep.*
- 3: /cgn2_6/ptodata/1/iaa/6A_COMB.pep.*
- 4: /cgn2_6/ptodata/1/iaa/6B_COMB.pep.*
- 5: /cgn2_6/ptodata/1/iaa/PCTUS_COMB.pep.*
- 6: /cgn2_6/ptodata/1/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	399	73.5	521	1	US-08-276-213-3
2	244.5	45.0	108	2	US-08-585-585A-2
3	244.5	45.0	108	2	US-08-249-037C-2
4	244.5	45.0	108	2	US-08-788-622B-2
5	244.5	45.0	108	3	US-08-788-621B-2
6	240.5	44.3	104	2	US-08-585-585A-3
7	240.5	44.3	104	2	US-08-249-037C-3
8	240.5	44.3	104	2	US-08-788-622B-3
9	240.5	44.3	104	3	US-08-788-621B-3
10	230.5	42.4	106	2	US-08-585-585A-4
11	230.5	42.4	106	2	US-08-249-037C-4
12	230.5	42.4	106	2	US-08-788-622B-4
13	230.5	42.4	106	3	US-08-788-621B-4
14	228.5	42.1	381	4	US-09-216-295-22
15	213.5	39.3	490	4	US-09-109-841-2
16	212.5	39.1	109	2	US-08-585-585A-1
17	212.5	39.1	109	2	US-08-249-037C-1
18	212.5	39.1	109	2	US-08-788-622B-1
19	212.5	39.1	109	3	US-08-788-621B-1
20	212.5	39.1	631	3	US-08-814-052-8
21	212.5	39.1	631	3	US-08-812-829-8
22	184.5	34.0	592	1	US-08-217-327-8
23	184	33.9	103	2	US-08-585-585A-5
24	184	33.9	103	2	US-08-249-037C-5
25	184	33.9	103	2	US-08-788-622B-5
26	184	33.9	103	3	US-08-788-621B-5
27	184	33.9	104	2	US-08-585-585A-6

28	184	33.9	104	2	US-08-249-037C-6	Sequence 6, Appl
29	184	33.9	104	2	US-08-788-622B-6	Sequence 6, Appl
30	184	33.9	104	3	US-08-788-621B-6	Sequence 6, Appl
31	179	33.0	430	2	US-08-924-440-2	Sequence 2, Appl
32	162.5	29.9	371	4	US-09-104-308-1	Sequence 1, Appl
33	162.5	29.9	371	4	US-09-321-981-1	Sequence 1, Appl
34	161.5	29.7	386	4	US-09-321-981-5	Sequence 5, Appl
35	122.5	22.6	97	2	US-08-585-585A-7	Sequence 7, Appl
36	122.5	22.6	97	2	US-08-249-037C-7	Sequence 7, Appl
37	122.5	22.6	97	2	US-08-788-622B-7	Sequence 7, Appl
38	122.5	22.6	97	3	US-08-788-621B-7	Sequence 7, Appl
39	116.5	21.5	427	2	US-08-282-197C-53	Sequence 53, Appl
40	94.5	17.4	312	4	US-09-216-295-21	Sequence 21, Appl
41	93	17.1	1026	1	US-08-194-290-7	Sequence 7, Appl
42	93	17.1	1026	2	US-08-614-377A-7	Sequence 7, Appl
43	93	17.1	1026	4	US-09-142-648B-7	Sequence 7, Appl
44	89.5	16.5	221	1	US-08-032-848C-13	Sequence 13, Appl
45	89.5	16.5	259	4	US-09-216-295-5	Sequence 5, Appl

ALIGNMENTS

RESULT 1
US-08-276-213-3
; Sequence 3, Application US/08276213
; Patent No. 5536655
; GENERAL INFORMATION:
; APPLICANT: Thomas, Steven
; APPLICANT: Laymon, Robert
; APPLICANT: Himmel, Michael
; TITLE OF INVENTION: GENE ENCODING FOR THE E1 ENDOGLUCANASE
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: National Renewable Energy Laboratory
; STREET: 1617 Cole Boulevard
; CITY: Golden
; STATE: CO
; COUNTRY: USA
; ZIP: 80401-3393
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.25
; CURRENT APPLICATION DATA: US/08/276,213
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: O'Connor, Edna
; REGISTRATION NUMBER: 29,252
; REFERENCE/DOCKET NUMBER: NREL IR# 94-08
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (303)231-1000
; TELEFAX: (303)231-1098
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 521 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHEetical: NO
; ANTI-SENSE: NO
; FRAGMENT TYPE: N-terminal
; US-08-276-213-3

Query Match 73.5% Score 399; DB 1; Length 521;
Best Local Similarity 69.3%; Pred. No. 1.5e-31;
Matches 70; Conservative 13; Mismatches 18; Indels 0; Gaps 0;

Oy 1 GVCCRTATVYVNDWGSFGTATVTVTCTSGTATVNTALTSQA 60

Db 421 GARTASYVNSDWGNGFTVTVAVTNSGVSATKTWTVSWTFGGNQTITNSWNAAVTQNG 480
QY 61 SVTATNLNYSNNVTOPGOSTTFEGNGSCGNTAPTLCCTAS 101
Db 481 SVTARNMSYNNVTOPGONTTFGQASVTGSGNAAPTACAAS 521

RESULT 2
US-08-585-585A-2
; Sequence 2, Application US/08585585A
; Patent No. 5874308
; GENERAL INFORMATION:
; APPLICANT: Kilburn, Douglas G.
; APPLICANT: Humphries, Keith R.
; APPLICANT: Doheny, James G.
; APPLICANT: Jervis, Eric
; APPLICANT: Allmonti, Jude
; TITLE OF INVENTION: Compositions and methods for modulating
; TITLE OF INVENTION: cell proliferation using growth factor-polysaccharide
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Rae-Venter Law Group
; STREET: Box 60039
; CITY: Palo Alto
; STATE: CA
; COUNTRY: U.S.
; ZIP: 94306
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/585,585A
; FILING DATE: 16-JAN-1996
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Kung, Viola T.
; REGISTRATION NUMBER: 41,131
; REFERENCE/DOCKET NUMBER: CBDT.016.000S
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (650)328-4400
; TELEFAX: (650)328-4477
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 108 amino acids
; TYPE: amino acid
; STRANDEDNESS: not relevant
; TOPOLOGY: not relevant
; MOLECULE TYPE: peptide
US-08-585-585A-2

Query Match 45.0%; Score 244.5; DB 2; Length 108;
Best Local Similarity 49.5%; Pred. No. 3.5e-17;
Matches 46; Conservative 14; Mismatches 32; Indels 1; Caps 1;
QY 3 GCRATVYVNSDWGSGFTATVTVTNTGSRATSGWTVVMSFGCGNQTIVNTWNTALTOSGASV 62
Db 5 GCOVLKGVN-QWNTGFTANTVTNTSSAPVDGWTLTFFSPSGQOVTOAWSSTVTQSGSAV 63
QY 63 TATNLSYNNVIOPGOSTTFEGNGSYGNTAPT 95
Db 64 TVRNAPWNGSIPAGGTAQFGFNGSHTGTNAAPT 96

RESULT 3
US-08-249-037C-2
; Sequence 2, Application US/08249037C
; Patent No. 5928917
; GENERAL INFORMATION:
; APPLICANT: Kilburn, Douglas G.

; APPLICANT: Miller, Robert C.
; APPLICANT: Warren, Richard A.J.
; APPLICANT: Gilkes, Neil R.
; TITLE OF INVENTION: Polysaccharide binding fusion proteins
; TITLE OF INVENTION: and conjugates
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Rae-Venter Law Group, P.C.
; STREET: P.O.Box 60039
; CITY: Palo Alto
; STATE: CA
; COUNTRY: U.S.
; ZIP: 94306
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/249,037C
; FILING DATE: 24-MAY-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/865,095
; FILING DATE: 08-APR-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/603,987
; FILING DATE: 25-OCT-1990
; APPLICATION NUMBER: US 07/216,794
; FILING DATE: 08-JUL-1988
; ATTORNEY/AGENT INFORMATION:
; NAME: Kung, Viola T.
; REGISTRATION NUMBER: 41,131
; REFERENCE/DOCKET NUMBER: CBDT.002.040S
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (650)328-4400
; TELEFAX: (650)328-4477
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 108 amino acids
; TYPE: amino acid
; STRANDEDNESS: not relevant
; TOPOLOGY: not relevant
; MOLECULE TYPE: peptide
US-08-249-037C-2

Query Match 45.0%; Score 244.5; DB 2; Length 108;
Best Local Similarity 49.5%; Pred. No. 3.5e-17;
Matches 46; Conservative 14; Mismatches 32; Indels 1; Caps 1;
QY 3 GCRATVYVNSDWGSGFTATVTVTNTGSRATSGWTVVMSFGCGNQTIVNTWNTALTOSGASV 62
Db 5 GCOVLKGVN-QWNTGFTANTVTNTSSAPVDGWTLTFFSPSGQOVTOAWSSTVTQSGSAV 63
QY 63 TATNLSYNNVIOPGOSTTFEGNGSYGNTAPT 95
Db 64 TVRNAPWNGSIPAGGTAQFGFNGSHTGTNAAPT 96

RESULT 4
US-08-788-622B-2
; Sequence 2, Application US/08788622B
; Patent No. 5962289
; GENERAL INFORMATION:
; APPLICANT: Kilburn, Douglas G.
; APPLICANT: Miller, Robert C.
; APPLICANT: Warren, Richard A.J.
; APPLICANT: Gilkes, Neil R.
; TITLE OF INVENTION: Polysaccharide binding fusion proteins
; TITLE OF INVENTION: and conjugates
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:

Query Match 44.3%; SCORE 240.5; DB 2; Length 104;
Best Local Similarity 45.5%; Pred. No. 8.3e-17;
Matches 46: Conservative 18; Mismatches 36; Indels 1

[illegible]

```

RESULT 8
US-08-788-622B-3
: Sequence 3, Application US/08788622B
: Patent No. 5962289
: GENERAL INFORMATION:
: APPLICANT: Kilburn, Douglas G.
: APPLICANT: Miller, Robert C.
: APPLICANT: Warren, Richard A.J.
: APPLICANT: Gilkes, Neil R.
: TITLE OF INVENTION: Polysaccharide binding fusion proteins
: TITLE OF INVENTION: and conjugates
: NUMBER OF SEQUENCES: 21
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Rae-Venter Law Group, P.C.
: STREET: P.O.Box 60039
: CITY: Palo Alto
: STATE: CA
: COUNTRY: U.S.
: ZIP: 94306
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC Compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08788,622B
: FILING DATE: January 23, 1997
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/249,037
: FILING DATE: 24-MAY-1994
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 07/865,095
: FILING DATE: 08-APR-1992
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 07/603,987
: FILING DATE: 25-OCT-1990
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 07/216,794
: FILING DATE: 08-JUL-1988
: ATTORNEY/AGENT INFORMATION:
: NAME: Kung, Viola T.
: REGISTRATION NUMBER: 41,131
: REFERENCE/DOCKET NUMBER: CBDT.002.06US
: TELECOMMUNICATION INFORMATION:

```


APPLICANT: Warren, Richard A.J.
APPLICANT: Gilkes, Neil R.
TITLE OF INVENTION: Polysaccharide binding fusion proteins
TITLE OF INVENTION: and conjugates
NUMBER OF SEQUENCES: 21
CORRESPONDENCE ADDRESS:
ADDRESSEE: Rae-Venter Law Group, P.C.
STREET: P.O.Box 60039
CITY: Palo Alto
STATE: CA
COUNTRY: U.S.
ZIP: 94306
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/788,621B
FILING DATE: January 23, 1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/249,037
FILING DATE: 24-MAY-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/865,095
FILING DATE: 08-APR-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/603,987
FILING DATE: 25-OCT-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/216,794
FILING DATE: 08-JUL-1988
ATTORNEY/AGENT INFORMATION:
NAME: Kung, Viola T.
REGISTRATION NUMBER: 41,131
REFERENCE/DOCKET NUMBER: CBDT.002.050S
TELECOMMUNICATION INFORMATION:
TELEPHONE: (650)328-4400
TELEFAX: (650)328-4477
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 106 amino acids
TYPE: amino acid
STRANDEDNESS: not relevant
TOPOLOGY: not relevant
MOLECULE TYPE: peptide
US-08-788-621B-4

Query Match 42.4%; Score 230.5; DB 3; Length 106;
Best Local Similarity 43.1%; Pred. No. 8.2e-16;
Matches 44; Conservative 18; Mismatches 33; Indels 7; Gaps 2;
Qy 4 CRATVVNSDMGSGFTATVTNTGSRATSGTWTVAWSFGNGQTVTNYNTALTOSGASVT 63
Db 4 CKVEYNASS-WNTGTASVRYTNTGTTALNGWTLTFEPANGQTVOQGWSADWSOGTGT 62
Qy 64 ATNLNNVNIOPQSGSTTFGNGSYSGTNTAPT-----LTCT 99
Db 63 AKNAWNGSLAAGTVDIGFNGAHNGTNNKPASFTLNCATCT 104

RESULT 14
US-09-216-295-22
Sequence 22, Application US/09216295
Patent No. 6268328
GENERAL INFORMATION:
APPLICANT: Mitchinson, Colin
APPLICANT: Wendt, Dan J.
TITLE OF INVENTION: No. 6268328el Variant EGIII-Like Cellulase Compositions
FILE REFERENCE: GC555
CURRENT APPLICATION NUMBER: US/09/216,295
CURRENT FILING DATE: 1998-12-18
NUMBER OF SEQ ID NOS: 41

SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 22
LENGTH: 381
TYPE: PRT
ORGANISM: Streptomyces lividans Celb
US-09-216-295-22

Query Match 42.1%; Score 228.5; DB 4; Length 381;
Best Local Similarity 48.9%; Pred. No. 5.8e-15;
Matches 45; Conservative 10; Mismatches 36; Indels 1; Gaps 1;
Qy 4 CRATVVNSDMGSGFTATVTNTGSRATSGTWTVAWSFGNGQTVTNYNTALTOSGASVT 63
Db 279 CAVSYGTAV-WODGFTADVTNTGTAPVDGWLAFITLPSGORITNAMNASLTPSSGVT 337
Qy 64 ATNLNNVNIOPQSGSTTFGNGSYSGTNTAPT 95
Db 338 ATGASHNARIAPGGSLSFGFGQTYGCAFAEPT 369

RESULT 15
US-09-109-841-2
Sequence 2, Application US/09109841
Patent No. 6207436
GENERAL INFORMATION:
APPLICANT: Bjornvad, Mads E.
APPLICANT: Hatakeyama, Mariko
APPLICANT: Schulein, Martin
APPLICANT: Nielsen, Jack B.
TITLE OF INVENTION: Endo-Beta-1, 4-Glucanases From
Saccharothrix
FILE REFERENCE: 5195.200-US
CURRENT APPLICATION NUMBER: US/09/109,841
CURRENT FILING DATE: 1998-07-02
EARLIER APPLICATION NUMBER: 0812/97
EARLIER FILING DATE: 1997-07-04
EARLIER APPLICATION NUMBER: 0846/97
EARLIER FILING DATE: 1997-07-11
EARLIER APPLICATION NUMBER: 60/053,506
EARLIER FILING DATE: 1997-07-23
NUMBER OF SEQ ID NOS: 4
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 2
LENGTH: 490
TYPE: PRT
ORGANISM: Saccharothrix australiensis
US-09-109-841-2

Query Match 39.3%; Score 213.5; DB 4; Length 490;
Best Local Similarity 44.1%; Pred. No. 2.3e-13;
Matches 41; Conservative 11; Mismatches 40; Indels 1; Gaps 1;
Qy 3 GCRATVVNSDMGSGFTATVTNTGSRATSGTWTVAWSFGNGQTVTNYNTALTOSGASV 62
Db 42 CCRVDYAVTSQMPGCGAAVVTNLGD-PLSSWELSWTFPPDCGVQOOLNNGVHSTSGSV 100
Qy 63 TATNLNNVNIOPQSGSTTFGNGSYSGTNTAPT 95
Db 101 TVKEMSWNGSVGTNASVQVGFNGSWNGANNAPT 133

Search completed: November 13, 2002, 11:56:06
Job time : 3.11997 secs

GenCore version 5.1.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 13, 2002, 11:55:11 : Search time 1.95905 Seconds
(without alignments)
776.473 Million cell updates/sec

Title: US-09-917-378-5
Perfect score: 543
Sequence: 1 GVGCRTATVYVNSDMGSGFTA.....GFNGSYSGTNTAPTTLCTAS 101

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 97044 seqs, 15060890 residues
Total number of hits satisfying chosen parameters: 97044

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_AA:*
1: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep.*
2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep.*
3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep.*
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
7: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep.*
8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep.*
10: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep.*
11: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep.*
12: /cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB.pep.*
13: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep.*
14: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	399	73.5	562	10	US-09-981-900B-5
2	162.5	29.9	371	10	US-09-739-861A-1
3	162.5	29.9	371	10	US-09-795-583-1
4	161.5	29.7	386	10	US-09-739-861A-5
5	161.5	29.7	386	10	US-09-795-583-5
6	93	17.1	1026	10	US-09-379-931-7
7	92	16.9	596	12	US-10-006-867-100
8	92	16.9	596	12	US-10-052-586-310
9	80.5	16.3	4440	12	US-10-052-586-310
10	86	15.8	344	10	US-09-770-621-2
11	83	15.3	1477	9	US-10-092-880-4
12	80.5	14.8	619	10	US-09-800-065-2
13	76	14.0	480	10	US-09-823-936-16
14	75.5	13.9	1435	12	US-10-052-586-581
15	73.5	13.5	688	10	US-09-864-761-36047
16	73	13.4	327	10	US-09-850-351A-23
17	73	13.4	327	10	US-09-850-351A-29
18	72.5	13.4	331	10	US-09-864-761-36048
19	72	13.3	1599	9	US-10-092-880-9

Sequence 10, Appl	9	US-10-092-880-10	1600	13.3	72	20
Sequence 1, Appl	1422	US-09-735-933-1	70.5	13.0	21	21
Sequence 2, Appl	4679	US-09-804-898-2	70.5	13.0	22	22
Sequence 34, Appl	218	US-09-916-494A-34	70	12.9	23	23
Sequence 3, Appl	234	US-10-075-872-3	70	12.9	24	24
Sequence 10, Appl	263	US-09-347-064-10	69.5	12.8	25	25
Sequence 4, Appl	267	US-09-347-064-4	69.5	12.8	26	26
Sequence 2, Appl	605	US-09-741-233A-2	69.5	12.8	27	27
Sequence 170, App	151	US-09-789-561-170	68.5	12.6	28	28
Sequence 131, App	473	US-09-910-059-131	68.5	12.6	29	29
Sequence 1065, Ap	957	US-09-922-217-1065	68.5	12.6	30	30
Sequence 1065, Ap	957	US-09-833-263-1065	68.5	12.6	31	31
Sequence 29, Appl	238	US-09-976-787-29	68	12.5	32	32
Sequence 28, Appl	238	US-09-865-198-28	68	12.5	33	33
Sequence 28, Appl	240	US-09-976-787-28	68	12.5	34	34
Sequence 27, Appl	240	US-09-865-198-27	68	12.5	35	35
Sequence 13542, A	472	US-09-815-242-13542	68	12.5	36	36
Sequence 36, Appl	479	US-09-815-242-13307	68	12.5	37	37
Sequence 4, Appl	525	US-09-804-551B-36	68	12.5	38	38
Sequence 37, Appl	234	US-10-075-872-4	67.5	12.4	39	39
Sequence 39, Appl	450	US-09-796-848A-37	67.5	12.4	40	40
Sequence 41, Appl	450	US-09-796-848A-39	67.5	12.4	41	41
Sequence 43, Appl	450	US-09-796-848A-41	67.5	12.4	42	42
Sequence 45, Appl	450	US-09-796-848A-43	67.5	12.4	43	43
Sequence 47, Appl	450	US-09-796-848A-45	67.5	12.4	44	44
	450	US-09-796-848A-47	67.5	12.4	45	45

ALIGNMENTS

RESULT 1
US-09-981-900B-5
; Sequence 5, Application US/09981900B
; Patent No. US20020138878A1
; GENERAL INFORMATION:
; APPLICANT: Sticklen, Masomeh B
; APPLICANT: Maqbool, Shahina B
; APPLICANT: Dale, Bruce E
; TITLE OF INVENTION: TRANSGENIC PLANTS CONTAINING LIGNINASE AND CELLULASE WHICH DEG
; FILE REFERENCE: MSU 4.1-539
; CURRENT APPLICATION NUMBER: US/09/981,900B
; PRIOR FILING DATE: 2002-03-18
; PRIOR APPLICATION NUMBER: 60/242,408
; PRIOR FILING DATE: 2000-10-20
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 5
; LENGTH: 562
; TYPE: PRT
; ORGANISM: Acidothermus cellulolyticus
US-09-981-900B-5

Query Match	73.5%	Score 399;	DB 10;	Length 562;
Best Local Similarity	69.3%	Pred. No. 3.7e-32;		
Matches	70;	Conservative 13;	Mismatches 18;	Indels 0; Caps 0;
OY	1	GVGCRTATVYVNSDMGSGFTATVTNTGSRATSGWTGVANSGGNQVTNTYNTALTQSGA	60	
Db	462	GARCTASYQVNSDMGNGFTVTVAVTNSGVSATKTWTVMSFTGCGNOTITNSNAAVTONGO	521	
QY	61	SVTATNLSYNNVLPQCGSTTFGFGNGSYSGTNTAPTTLCTAS	101	
Db	522	SVTARNNSYNNVLPQCGNTTFGFGASVTGSNAAPTACAAAS	562	
RESULT 2				
US-09-739-861A-1				
; Sequence 1, Application US/09739861A				
; Patent No. US20020076792A1				
; GENERAL INFORMATION:				
; APPLICANT: Jones, Brian E.				


```
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 5
; LENGTH: 386
; TYPE: PRT
; ORGANISM: Unknown
; FEATURE:
; OTHER INFORMATION: Nearest "neighbor" = Streptomyces
; OTHER INFORMATION: thermovibaceus
US-09-795-583-5

Query Match          29.7%  Score 161.5;  DB 10;  Length 386;
Best Local Similarity 33.7%;  Pred. No. 6.2e-09;
Matches 35;  Conservative 19;  Mismatches 41;  Indels 9;  Gaps 4;

QY 4 CRATVNVNSDMGSGFTATVTNTGSRATSGWTVAWSFGGQQTVTNYNNTALTQSGASVT 63
   I : I : I : I : I : I : I : I : I : I : I : I : I : I : I : I :
Db 282 COVSYSHTWPGGFTVDTTITNGSTPVDGWELDFTLPGAGTIVTSANWALLISPASGAVT 340
   I : I : I : I : I : I : I : I : I : I : I : I : I : I : I : I :

QY 64 ATNLSYN-NVTPQGOSTTFGNGSYSGTN-TAPT-----LTCT 99
   I : I : I : I : I : I : I : I : I : I : I : I : I : I : I : I :
Db 341 ARSTGSGNRJAANGTOSFGQGTQCTSSGTGFENAPAGGRLLNGTSTCT 384
   I : I : I : I : I : I : I : I : I : I : I : I : I : I : I : I :

RESULT 6
US-09-379-931-7
; Sequence 7, Application US/09379931
; Patent No. US20020009792A1
; GENERAL INFORMATION:
; APPLICANT: Smlt, John
; APPLICANT: Bingle, Wade H.
; TITLE OF INVENTION: EXPRESSION AND SECRETION OF HETEROLOGOUS POLYPEPTIDES FROM CAULC
; FILE REFERENCE: 08106/002003
; CURRENT APPLICATION NUMBER: US/09/379,931
; CURRENT FILING DATE: 1999-08-24
; PRIOR APPLICATION NUMBER: US 08/614,377
; PRIOR FILING DATE: 1996-03-12
; PRIOR APPLICATION NUMBER: US 08/194,290
; PRIOR FILING DATE: 1994-02-09
; PRIOR APPLICATION NUMBER: US 07/895,367
; PRIOR FILING DATE: 1992-06-09
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 7
; LENGTH: 1026
; TYPE: PRT
; ORGANISM: Caulobacter crescentus
US-09-379-931-7

Query Match          17.1%  Score 93;  DB 10;  Length 1026;
Best Local Similarity 34.0%;  Pred. No. 0.1;
Matches 32;  Conservative 11;  Mismatches 39;  Indels 12;  Gaps 3;

QY 15 GSGFTATVTNTGSRATSGWTVAWSFGGQQTVTNYNNTALTQSGASVTATNLSTY----- 69
   I : I : I : I : I : I : I : I : I : I : I : I : I : I : I : I :
Db 450 GNSSTTATVTQTAA-ATAGATVAGRVNGAVTITDASAASATTAGKIATVTLGSGAATI 508
   I : I : I : I : I : I : I : I : I : I : I : I : I : I : I : I :

QY 70 -----NNVIQPGQSTTFGCF-NGSYSGTNTAPTTLT 97
   I : I : I : I : I : I : I : I : I : I : I : I : I : I : I : I :
Db 509 DSSALTTVNLSTGCTSLGIGRGALTAPTANTLT 542
   I : I : I : I : I : I : I : I : I : I : I : I : I : I : I : I :

RESULT 7
US-10-006-867-100
; Sequence 100, Application US/10006867
; Patent No. US20020119130A1
; GENERAL INFORMATION:
; APPLICANT: Eaton, Dan L.
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Gerlitsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Grimaldi, Christopher J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Watanabe, Colin K.
; APPLICANT: Wood, William I.
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3230R1C1
; CURRENT APPLICATION NUMBER: US/10/006,867
; CURRENT FILING DATE: 2001-12-06
; PRIOR APPLICATION NUMBER: 60/063435
; PRIOR FILING DATE: 1997-10-29
; PRIOR APPLICATION NUMBER: 60/064215
; PRIOR FILING DATE: 1997-10-29
; PRIOR APPLICATION NUMBER: 60/082797
; PRIOR FILING DATE: 1998-04-22
; PRIOR APPLICATION NUMBER: 60/083495
; PRIOR FILING DATE: 1998-04-29
; PRIOR APPLICATION NUMBER: 60/085579
; PRIOR FILING DATE: 1998-05-15
; PRIOR APPLICATION NUMBER: 60/087759
; PRIOR FILING DATE: 1998-06-02
; PRIOR APPLICATION NUMBER: 60/088021
; PRIOR FILING DATE: 1998-06-04
; PRIOR APPLICATION NUMBER: 60/088029
; PRIOR FILING DATE: 1998-06-04
; PRIOR APPLICATION NUMBER: 60/088030
; PRIOR FILING DATE: 1998-06-04
; PRIOR APPLICATION NUMBER: 60/088734
; PRIOR FILING DATE: 1998-06-10
; PRIOR APPLICATION NUMBER: 60/088740
; PRIOR FILING DATE: 1998-06-10
; PRIOR APPLICATION NUMBER: 60/088811
; PRIOR FILING DATE: 1998-06-10
; PRIOR APPLICATION NUMBER: 60/088824
; PRIOR FILING DATE: 1998-06-10
; PRIOR APPLICATION NUMBER: 60/088825
; PRIOR FILING DATE: 1998-06-10
; PRIOR APPLICATION NUMBER: 60/088863
; PRIOR FILING DATE: 1998-06-11
; PRIOR APPLICATION NUMBER: 60/089105
; PRIOR FILING DATE: 1998-06-12
; PRIOR APPLICATION NUMBER: 60/089514
; PRIOR FILING DATE: 1998-06-16
; PRIOR APPLICATION NUMBER: 60/089653
; PRIOR FILING DATE: 1998-06-17
; PRIOR APPLICATION NUMBER: 60/089952
; PRIOR FILING DATE: 1998-06-19
; PRIOR APPLICATION NUMBER: 60/090246
; PRIOR FILING DATE: 1998-06-22
; PRIOR APPLICATION NUMBER: 60/090444
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090688
; PRIOR FILING DATE: 1998-06-25
; PRIOR APPLICATION NUMBER: 60/090696
; PRIOR FILING DATE: 1998-06-25
; PRIOR APPLICATION NUMBER: 60/090862
; PRIOR FILING DATE: 1998-06-26
; PRIOR APPLICATION NUMBER: 60/091628
; PRIOR FILING DATE: 1998-07-02
; PRIOR APPLICATION NUMBER: 60/096012
; PRIOR FILING DATE: 1998-08-10
; PRIOR APPLICATION NUMBER: 60/096757
; PRIOR FILING DATE: 1998-08-17
; PRIOR APPLICATION NUMBER: 60/096949
; PRIOR FILING DATE: 1998-08-18
; PRIOR APPLICATION NUMBER: 60/096959
; PRIOR FILING DATE: 1998-08-18
; PRIOR APPLICATION NUMBER: 60/097954
; PRIOR FILING DATE: 1998-08-26
; PRIOR APPLICATION NUMBER: 60/097971
; PRIOR FILING DATE: 1998-08-26
; PRIOR APPLICATION NUMBER: 60/097979
; PRIOR FILING DATE: 1998-08-26
```

; PRIOR APPLICATION NUMBER: 60/098749
; PRIOR FILING DATE: 1998-09-01
; PRIOR APPLICATION NUMBER: 60/099741
; PRIOR FILING DATE: 1998-09-10
; PRIOR APPLICATION NUMBER: 60/099763
; PRIOR FILING DATE: 1998-09-10
; PRIOR APPLICATION NUMBER: 60/099792
; PRIOR FILING DATE: 1998-09-10
; PRIOR APPLICATION NUMBER: 60/099812
; PRIOR FILING DATE: 1998-09-10
; PRIOR APPLICATION NUMBER: 60/099815
; PRIOR FILING DATE: 1998-09-10
; PRIOR APPLICATION NUMBER: 60/100627
; PRIOR FILING DATE: 1998-09-16
; PRIOR APPLICATION NUMBER: 60/100662
; PRIOR FILING DATE: 1998-09-16
; PRIOR APPLICATION NUMBER: 60/100683
; PRIOR FILING DATE: 1998-09-17
; PRIOR APPLICATION NUMBER: 60/100684
; PRIOR FILING DATE: 1998-09-17
; PRIOR APPLICATION NUMBER: 60/100930
; PRIOR FILING DATE: 1998-09-17
; PRIOR APPLICATION NUMBER: 60/101279
; PRIOR FILING DATE: 1998-09-22
; PRIOR APPLICATION NUMBER: 60/101475
; PRIOR FILING DATE: 1998-09-23
; PRIOR APPLICATION NUMBER: 60/101738
; PRIOR FILING DATE: 1998-09-24
; PRIOR APPLICATION NUMBER: 60/101743
; PRIOR FILING DATE: 1998-09-24
; PRIOR APPLICATION NUMBER: 60/101916
; PRIOR FILING DATE: 1998-09-24
; PRIOR APPLICATION NUMBER: 60/102570
; PRIOR FILING DATE: 1998-09-30
; PRIOR APPLICATION NUMBER: 60/103449
; PRIOR FILING DATE: 1998-10-06
; PRIOR APPLICATION NUMBER: 60/103678
; PRIOR FILING DATE: 1998-10-08
; PRIOR APPLICATION NUMBER: 60/103679
; PRIOR FILING DATE: 1998-10-08
; PRIOR APPLICATION NUMBER: 60/103711
; PRIOR FILING DATE: 1998-10-08
; PRIOR APPLICATION NUMBER: 60/105000
; PRIOR FILING DATE: 1998-10-20
; PRIOR APPLICATION NUMBER: 60/105002
; PRIOR FILING DATE: 1998-10-20
; PRIOR APPLICATION NUMBER: 60/105881
; PRIOR FILING DATE: 1998-10-27
; PRIOR APPLICATION NUMBER: 60/106030
; PRIOR FILING DATE: 1998-10-28
; PRIOR APPLICATION NUMBER: 60/106464
; PRIOR FILING DATE: 1998-10-30
; PRIOR APPLICATION NUMBER: 60/106856
; PRIOR FILING DATE: 1998-11-03
; PRIOR APPLICATION NUMBER: 60/108807
; PRIOR FILING DATE: 1998-11-17
; PRIOR APPLICATION NUMBER: 60/112419
; PRIOR FILING DATE: 1998-12-15
; PRIOR APPLICATION NUMBER: 60/112422
; PRIOR FILING DATE: 1998-12-15
; PRIOR APPLICATION NUMBER: 60/112853
; PRIOR FILING DATE: 1998-12-16
; PRIOR APPLICATION NUMBER: 60/113011
; PRIOR FILING DATE: 1998-12-16
; PRIOR APPLICATION NUMBER: 60/112854
; PRIOR FILING DATE: 1998-12-16
; PRIOR APPLICATION NUMBER: 60/113300
; PRIOR FILING DATE: 1998-12-22
; PRIOR APPLICATION NUMBER: 60/113408
; PRIOR FILING DATE: 1998-12-22
; PRIOR APPLICATION NUMBER: 60/113430
; PRIOR FILING DATE: 1998-12-23
; PRIOR APPLICATION NUMBER: 60/113621

; PRIOR FILING DATE: 1998-12-23
; PRIOR APPLICATION NUMBER: 60/114223
; PRIOR FILING DATE: 1998-12-30
; PRIOR APPLICATION NUMBER: 60/115614
; PRIOR FILING DATE: 1999-01-12
; PRIOR APPLICATION NUMBER: 60/116527
; PRIOR FILING DATE: 1999-01-20
; PRIOR APPLICATION NUMBER: 60/116843
; PRIOR FILING DATE: 1999-01-22
; PRIOR APPLICATION NUMBER: 60/119285
; PRIOR FILING DATE: 1999-02-09
; PRIOR APPLICATION NUMBER: 60/119287
; PRIOR FILING DATE: 1999-02-09
; PRIOR APPLICATION NUMBER: 60/119525
; PRIOR FILING DATE: 1999-02-10
; PRIOR APPLICATION NUMBER: 60/119549
; PRIOR FILING DATE: 1999-02-10
; PRIOR APPLICATION NUMBER: 60/120014
; PRIOR FILING DATE: 1999-02-11
; PRIOR APPLICATION NUMBER: 60/129122
; PRIOR FILING DATE: 1999-04-13
; PRIOR APPLICATION NUMBER: 60/129674
; PRIOR FILING DATE: 1999-04-16
; PRIOR APPLICATION NUMBER: 60/131291
; PRIOR FILING DATE: 1999-04-27
; PRIOR APPLICATION NUMBER: 60/138387
; PRIOR FILING DATE: 1999-06-09
; PRIOR APPLICATION NUMBER: 60/144791
; PRIOR FILING DATE: 1999-07-20
; PRIOR APPLICATION NUMBER: 60/169495
; PRIOR FILING DATE: 1999-12-07
; PRIOR APPLICATION NUMBER: 60/175481
; PRIOR FILING DATE: 2000-01-11
; PRIOR APPLICATION NUMBER: 60/191007
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/199397
; PRIOR FILING DATE: 2000-04-25
; PRIOR APPLICATION NUMBER: 09/380139
; PRIOR FILING DATE: 1998-08-25
; PRIOR APPLICATION NUMBER: 09/311832
; PRIOR FILING DATE: 1999-05-14
; PRIOR APPLICATION NUMBER: 09/380137
; PRIOR FILING DATE: 1999-08-25
; PRIOR APPLICATION NUMBER: 09/380138
; PRIOR FILING DATE: 1999-08-25
; PRIOR APPLICATION NUMBER: 09/380142

Query Match 16.9%; Score 92; DB 12; Length 596;

Best Local Similarity 35.0%; Pred. No. 0.071;

Matches 35; Conservative 12; Mismatches 39; Indels 14; Gaps 5;

Qy 2 VGRATYVNVSDMGSGFTATVTVTGSRATSGWTVMVSGGNOTVTNYWNTALTQSGAS, 61

Db 218 VSSRATATNSESSTTSSGASTATNSESRTTS-----NGAGTATNS-ESSTTSSGAS 268

Qy 62 VTATNLSYNNVIOPGOSTTGFENGSGTNTAPTTLTCTAS 101

Db 269 -TATN-SDSSTVSSGASTA---TNSSESSTTSSGASTATNS 303

RESULT 8

US-10-052-586-310

; Sequence 310, Application US/10052586

; Patent No. US20020127584A1

; GENERAL INFORMATION:

; APPLICANT: Baker, Kevin P.

; APPLICANT: Chen, Jian

; APPLICANT: Desnoyers, Luc

; APPLICANT: Goddard, Audrey

; APPLICANT: Godowski, Paul J.

; APPLICANT: Gurney, Austin L.

; APPLICANT: Pan, James

; APPLICANT: Smith, Victoria

APPLICANT: Watanabe, Collin K.
APPLICANT: Wood, William I.
APPLICANT: Zhang, Zemin
TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
FILE REFERENCE: P3430R1C1
CURRENT APPLICATION NUMBER: US/10/052,586
CURRENT FILING DATE: 2002-01-15
PRIOR APPLICATION NUMBER: 60/059263
PRIOR FILING DATE: 1997-09-18
PRIOR APPLICATION NUMBER: 60/059266
PRIOR FILING DATE: 1997-09-18
PRIOR APPLICATION NUMBER: 60/062250
PRIOR FILING DATE: 1997-10-17
PRIOR APPLICATION NUMBER: 60/063120
PRIOR FILING DATE: 1997-10-24
PRIOR APPLICATION NUMBER: 60/063121
PRIOR FILING DATE: 1997-10-24
PRIOR APPLICATION NUMBER: 60/063486
PRIOR FILING DATE: 1997-10-21
PRIOR APPLICATION NUMBER: 60/063540
PRIOR FILING DATE: 1997-10-28
PRIOR APPLICATION NUMBER: 60/063541
PRIOR FILING DATE: 1997-10-28
PRIOR APPLICATION NUMBER: 60/063544
PRIOR FILING DATE: 1997-10-28
PRIOR APPLICATION NUMBER: 60/063564
PRIOR FILING DATE: 1997-10-28
PRIOR APPLICATION NUMBER: 60/063734
PRIOR FILING DATE: 1997-10-29
PRIOR APPLICATION NUMBER: 60/063870
PRIOR FILING DATE: 1997-10-31
PRIOR APPLICATION NUMBER: 60/064103
PRIOR FILING DATE: 1997-10-31
PRIOR APPLICATION NUMBER: 60/065311
PRIOR FILING DATE: 1997-11-13
PRIOR APPLICATION NUMBER: 60/066120
PRIOR FILING DATE: 1997-11-21
PRIOR APPLICATION NUMBER: 60/066466
PRIOR FILING DATE: 1997-11-24
PRIOR APPLICATION NUMBER: 60/066772
PRIOR FILING DATE: 1997-11-24
PRIOR APPLICATION NUMBER: 60/069335
PRIOR FILING DATE: 1997-12-11
PRIOR APPLICATION NUMBER: 60/069425
PRIOR FILING DATE: 1997-12-12
PRIOR APPLICATION NUMBER: 60/069870
PRIOR FILING DATE: 1997-12-17
PRIOR APPLICATION NUMBER: 60/068017
PRIOR FILING DATE: 1997-12-18
PRIOR APPLICATION NUMBER: 60/077450
PRIOR FILING DATE: 1998-03-10
PRIOR APPLICATION NUMBER: 60/077632
PRIOR FILING DATE: 1998-03-11
PRIOR APPLICATION NUMBER: 60/077649
PRIOR FILING DATE: 1998-03-11
PRIOR APPLICATION NUMBER: 60/078886
PRIOR FILING DATE: 1998-03-20
PRIOR APPLICATION NUMBER: 60/078939
PRIOR FILING DATE: 1998-03-20
PRIOR APPLICATION NUMBER: 60/079664
PRIOR FILING DATE: 1998-03-27
PRIOR APPLICATION NUMBER: 60/079786
PRIOR FILING DATE: 1998-03-27
PRIOR APPLICATION NUMBER: 60/080107
PRIOR FILING DATE: 1998-03-31
PRIOR APPLICATION NUMBER: 60/080194
PRIOR FILING DATE: 1998-03-31
PRIOR APPLICATION NUMBER: 60/080327
PRIOR FILING DATE: 1998-04-01
PRIOR APPLICATION NUMBER: 60/080333
PRIOR FILING DATE: 1998-04-01
PRIOR APPLICATION NUMBER: 60/081049
PRIOR FILING DATE: 1998-04-08
PRIOR APPLICATION NUMBER: 60/081070
PRIOR FILING DATE: 1998-04-08
PRIOR APPLICATION NUMBER: 60/081195
PRIOR FILING DATE: 1998-04-09
PRIOR APPLICATION NUMBER: 60/081838
PRIOR FILING DATE: 1998-04-15
PRIOR APPLICATION NUMBER: 60/082568
PRIOR FILING DATE: 1998-04-21
PRIOR APPLICATION NUMBER: 60/082569
PRIOR FILING DATE: 1998-04-21
PRIOR APPLICATION NUMBER: 60/082704
PRIOR FILING DATE: 1998-04-22
PRIOR APPLICATION NUMBER: 60/082797
PRIOR FILING DATE: 1998-04-22
PRIOR APPLICATION NUMBER: 60/083322
PRIOR FILING DATE: 1998-04-28
PRIOR APPLICATION NUMBER: 60/083495
PRIOR FILING DATE: 1998-04-29
PRIOR APPLICATION NUMBER: 60/083496
PRIOR FILING DATE: 1998-04-29
PRIOR APPLICATION NUMBER: 60/083499
PRIOR FILING DATE: 1998-04-29
PRIOR APPLICATION NUMBER: 60/083559
PRIOR FILING DATE: 1998-04-29
PRIOR APPLICATION NUMBER: 60/084366
PRIOR FILING DATE: 1998-05-05
PRIOR APPLICATION NUMBER: 60/084414
PRIOR FILING DATE: 1998-05-06
PRIOR APPLICATION NUMBER: 60/084639
PRIOR FILING DATE: 1998-05-07
PRIOR APPLICATION NUMBER: 60/084640
PRIOR FILING DATE: 1998-05-07
PRIOR APPLICATION NUMBER: 60/084643
PRIOR FILING DATE: 1998-05-07
PRIOR APPLICATION NUMBER: 60/085573
PRIOR FILING DATE: 1998-05-15
PRIOR APPLICATION NUMBER: 60/085579
PRIOR FILING DATE: 1998-05-15
PRIOR APPLICATION NUMBER: 60/085580
PRIOR FILING DATE: 1998-05-15
PRIOR APPLICATION NUMBER: 60/085582
PRIOR FILING DATE: 1998-05-15
PRIOR APPLICATION NUMBER: 60/085700
PRIOR FILING DATE: 1998-05-15
PRIOR APPLICATION NUMBER: 60/086023
PRIOR FILING DATE: 1998-05-18
PRIOR APPLICATION NUMBER: 60/086392
PRIOR FILING DATE: 1998-05-22
PRIOR APPLICATION NUMBER: 60/086486
PRIOR FILING DATE: 1998-05-22
PRIOR APPLICATION NUMBER: 60/087098
PRIOR FILING DATE: 1998-05-28
PRIOR APPLICATION NUMBER: 60/087208
PRIOR FILING DATE: 1998-05-28
PRIOR APPLICATION NUMBER: 60/087609
PRIOR FILING DATE: 1998-06-02
PRIOR APPLICATION NUMBER: 60/087759
PRIOR FILING DATE: 1998-06-02
PRIOR APPLICATION NUMBER: 60/087827
PRIOR FILING DATE: 1998-06-03
PRIOR APPLICATION NUMBER: 60/088025
PRIOR FILING DATE: 1998-06-04
PRIOR APPLICATION NUMBER: 60/088028
PRIOR FILING DATE: 1998-06-04
PRIOR APPLICATION NUMBER: 60/088029
PRIOR FILING DATE: 1998-06-04
PRIOR APPLICATION NUMBER: 60/088033
PRIOR FILING DATE: 1998-06-04
PRIOR APPLICATION NUMBER: 60/088167
PRIOR FILING DATE: 1998-06-05
PRIOR APPLICATION NUMBER: 60/088202
PRIOR FILING DATE: 1998-06-05

```

; PRIOR APPLICATION NUMBER: 60/088212
; PRIOR FILING DATE: 1998-06-05
; PRIOR APPLICATION NUMBER: 60/088217
; PRIOR FILING DATE: 1998-06-05
; PRIOR APPLICATION NUMBER: 60/088326
; PRIOR FILING DATE: 1998-06-04
; PRIOR APPLICATION NUMBER: 60/088655
; PRIOR FILING DATE: 1998-06-09
; PRIOR APPLICATION NUMBER: 60/088722
; PRIOR FILING DATE: 1998-06-10
; PRIOR APPLICATION NUMBER: 60/088738
; PRIOR FILING DATE: 1998-06-10
; PRIOR APPLICATION NUMBER: 60/088740
; PRIOR FILING DATE: 1998-06-10
; PRIOR APPLICATION NUMBER: 60/088811
; PRIOR FILING DATE: 1998-06-10
; PRIOR APPLICATION NUMBER: 60/088824
; PRIOR FILING DATE: 1998-06-10
; PRIOR APPLICATION NUMBER: 60/088825
; PRIOR FILING DATE: 1998-06-10
; PRIOR APPLICATION NUMBER: 60/088826
; PRIOR FILING DATE: 1998-06-10
; PRIOR APPLICATION NUMBER: 60/088861
; PRIOR FILING DATE: 1998-06-11
; PRIOR APPLICATION NUMBER: 60/088863
; PRIOR FILING DATE: 1998-06-11
; PRIOR APPLICATION NUMBER: 60/088876
; PRIOR FILING DATE: 1998-06-11
; PRIOR APPLICATION NUMBER: 60/089090
; PRIOR FILING DATE: 1998-06-12
; PRIOR APPLICATION NUMBER: 60/089105
; PRIOR FILING DATE: 1998-06-12
; PRIOR APPLICATION NUMBER: 60/089512
; PRIOR FILING DATE: 1998-06-16
; PRIOR APPLICATION NUMBER: 60/089514
; PRIOR FILING DATE: 1998-06-16
; PRIOR APPLICATION NUMBER: 60/089538
; PRIOR FILING DATE: 1998-06-17
; PRIOR APPLICATION NUMBER: 60/089598
; PRIOR FILING DATE: 1998-06-17
; PRIOR APPLICATION NUMBER: 60/089653
; PRIOR FILING DATE: 1998-06-17
; PRIOR APPLICATION NUMBER: 60/089908
; PRIOR FILING DATE: 1998-06-17
;
Query Match          16.9%  Score 92:  DB 12:  Length 596:
Best Local Similarity 35.0%  Pred. No. 0.071:
Matches 35:  Conservative 12:  Mismatches 39:  Indels 14:  Gaps 5:
QY  2  VCKRATVYVNSDMGSGFTATVTVTNTGSRATSGWTVAWFGGNGQTVTNYWNTALTOSGAS 61
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db  218  VSSRSTATNSESTTSSGASTATNSESRRTS-----NGAGTATNS-ESSTTSSGAS 268
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY  62  VTATNLNYYNVIOPGOSTTGFNGSYSGTNTAPTTLCTAS 101
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db  269  -TATN-SDSSTVSSGASTA---TNSESSTTSSGASTATNS 303
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
RESULT 9
US-10-052-586-525
; Sequence 525, Application US/10052586
; Patent No. US20020127584A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Chen, Jian
; APPLICANT: Desnoyers, Luc
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Pan, James
; APPLICANT: Smith, Victoria
; APPLICANT: Watanabe, Colin K.
; APPLICANT: Wood, William I.
; APPLICANT: Zhang, Zemin
;
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3430RIC1
; CURRENT APPLICATION NUMBER: US/10/052.586
; CURRENT FILING DATE: 2002-01-15
; PRIOR APPLICATION NUMBER: 60/059263
; PRIOR FILING DATE: 1997-09-18
; PRIOR APPLICATION NUMBER: 60/059266
; PRIOR FILING DATE: 1997-09-18
; PRIOR APPLICATION NUMBER: 60/062250
; PRIOR FILING DATE: 1997-10-17
; PRIOR APPLICATION NUMBER: 60/063120
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: 60/063121
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: 60/063486
; PRIOR FILING DATE: 1997-10-21
; PRIOR APPLICATION NUMBER: 60/063540
; PRIOR FILING DATE: 1997-10-28
; PRIOR APPLICATION NUMBER: 60/063541
; PRIOR FILING DATE: 1997-10-28
; PRIOR APPLICATION NUMBER: 60/063544
; PRIOR FILING DATE: 1997-10-28
; PRIOR APPLICATION NUMBER: 60/063564
; PRIOR FILING DATE: 1997-10-28
; PRIOR APPLICATION NUMBER: 60/063734
; PRIOR FILING DATE: 1997-10-29
; PRIOR APPLICATION NUMBER: 60/063870
; PRIOR FILING DATE: 1997-10-31
; PRIOR APPLICATION NUMBER: 60/064103
; PRIOR FILING DATE: 1997-10-31
; PRIOR APPLICATION NUMBER: 60/065311
; PRIOR FILING DATE: 1997-11-13
; PRIOR APPLICATION NUMBER: 60/066120
; PRIOR FILING DATE: 1997-11-21
; PRIOR APPLICATION NUMBER: 60/066466
; PRIOR FILING DATE: 1997-11-24
; PRIOR APPLICATION NUMBER: 60/066772
; PRIOR FILING DATE: 1997-11-24
; PRIOR APPLICATION NUMBER: 60/069335
; PRIOR FILING DATE: 1997-12-11
; PRIOR APPLICATION NUMBER: 60/069425
; PRIOR FILING DATE: 1997-12-12
; PRIOR APPLICATION NUMBER: 60/069870
; PRIOR FILING DATE: 1997-12-17
; PRIOR APPLICATION NUMBER: 60/068017
; PRIOR FILING DATE: 1997-12-18
; PRIOR APPLICATION NUMBER: 60/077450
; PRIOR FILING DATE: 1998-03-10
; PRIOR APPLICATION NUMBER: 60/077632
; PRIOR FILING DATE: 1998-03-11
; PRIOR APPLICATION NUMBER: 60/077649
; PRIOR FILING DATE: 1998-03-11
; PRIOR APPLICATION NUMBER: 60/078886
; PRIOR FILING DATE: 1998-03-20
; PRIOR APPLICATION NUMBER: 60/078939
; PRIOR FILING DATE: 1998-03-20
; PRIOR APPLICATION NUMBER: 60/079664
; PRIOR FILING DATE: 1998-03-27
; PRIOR APPLICATION NUMBER: 60/079786
; PRIOR FILING DATE: 1998-03-27
; PRIOR APPLICATION NUMBER: 60/080107
; PRIOR FILING DATE: 1998-03-31
; PRIOR APPLICATION NUMBER: 60/080194
; PRIOR FILING DATE: 1998-03-31
; PRIOR APPLICATION NUMBER: 60/080327
; PRIOR FILING DATE: 1998-04-01
; PRIOR APPLICATION NUMBER: 60/080333
; PRIOR FILING DATE: 1998-04-01
; PRIOR APPLICATION NUMBER: 60/081049
; PRIOR FILING DATE: 1998-04-08
; PRIOR APPLICATION NUMBER: 60/081070
; PRIOR FILING DATE: 1998-04-08
;

```

1	PRIOR FILING DATE: 1998-06-05
2	PRIOR APPLICATION NUMBER: 60/088336
3	PRIOR FILING DATE: 1998-06-04
4	PRIOR APPLICATION NUMBER: 60/088655
5	PRIOR FILING DATE: 1998-06-09
6	PRIOR APPLICATION NUMBER: 60/088722
7	PRIOR FILING DATE: 1998-06-10
8	PRIOR APPLICATION NUMBER: 60/088738
9	PRIOR FILING DATE: 1998-06-10
10	PRIOR APPLICATION NUMBER: 60/088740
11	PRIOR FILING DATE: 1998-06-10
12	PRIOR APPLICATION NUMBER: 60/088811
13	PRIOR FILING DATE: 1998-06-10
14	PRIOR APPLICATION NUMBER: 60/088824
15	PRIOR FILING DATE: 1998-06-10
16	PRIOR APPLICATION NUMBER: 60/088825
17	PRIOR FILING DATE: 1998-06-10
18	PRIOR APPLICATION NUMBER: 60/088826
19	PRIOR FILING DATE: 1998-06-10
20	PRIOR APPLICATION NUMBER: 60/088861
21	PRIOR FILING DATE: 1998-06-11
22	PRIOR APPLICATION NUMBER: 60/088863
23	PRIOR FILING DATE: 1998-06-11
24	PRIOR APPLICATION NUMBER: 60/088876
25	PRIOR FILING DATE: 1998-06-11
26	PRIOR APPLICATION NUMBER: 60/089090
27	PRIOR FILING DATE: 1998-06-12
28	PRIOR APPLICATION NUMBER: 60/089105
29	PRIOR FILING DATE: 1998-06-12
30	PRIOR APPLICATION NUMBER: 60/089512
31	PRIOR FILING DATE: 1998-06-16
32	PRIOR APPLICATION NUMBER: 60/089514
33	PRIOR FILING DATE: 1998-06-16
34	PRIOR APPLICATION NUMBER: 60/089538
35	PRIOR FILING DATE: 1998-06-17
36	PRIOR APPLICATION NUMBER: 60/089598
37	PRIOR FILING DATE: 1998-06-17
38	PRIOR APPLICATION NUMBER: 60/089653
39	PRIOR FILING DATE: 1998-06-17
40	PRIOR APPLICATION NUMBER: 60/089908

Qy 3 GCRATYYVNSDVGSGFTATVTNTGSRATSGWTVANSEF-----GNQTIV 47
- - - : - : - - - - - : - - - - -
Db 4094 GAATCTTTCACGTAGACACTATTATTTGGCGTTT--TCACATGTCCTGCAAATGTATGTGTTAA 415

RESULT 10

? GENNAME INFORMATION:
 ? APPLICANT: M ntyl , Arja
 ? APPLICANT: Vehmaanper , Jari
 ? APPLICANT: Fagerstr m, Richard
 ? APPLICANT: Lantto, Raija
 ? APPLICANT: Paloheimo, Marja
 ? APPLICANT: Suominen, Pirkko
 ? APPLICANT: Lahtinen, Tarja
 ? TITLE OF INVENTION: Production and Secretion of Proteins of
 ? NUMBER OF SEQUENCES: 39
 ? CORRESPONDENCE ADDRESS:
 ? ADDRESSEE: STERNE, KESSLER, GOLDSTEIN & FOX, P.L.L.C.
 ? STREET: 1100 New York Ave., N.W. Suite 600
 ? CITY: Washington
 ? STATE: D.C.

```

COUNTRY: U.S.A.
ZIP: 20005
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, version #1.30
CURRENT APPLICATION DATA: US/09/770,621
APPLICATION NUMBER: US/09/770,621
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/590,563
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/332,412
FILING DATE: 31-OCT-1994
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/282,001
FILING DATE: 29-JUL-1994
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Bugalsky, Lawrence B.
REGISTRATION NUMBER: 35,086
REFERENCE/DOCKET NUMBER: 1050.0340003
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-371-2600
TELEFAX: 202-371-2540
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 344 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-770-621-2

Query Match 15.8%; Score 86; DB 10; Length 344;
Best Local Similarity 24.5%; Pred. No. 0.15;
Matches 25; Conservative 14; Mismatches 49; Indels 14; Gaps 3;

QY 1 GVGRATVYNVDGSGFTATVTNTGSRATSGTWTVAWSFGGQTVTNYNNTALTQSCA 60
   | | | | | : | | : | | | | : | |
Db 256 GCGCATLISAGQQSDRYNLNVSVS-----GSNNWTVRMDVPYPARIITWNIIHQWPES 310
   | : | | : | | : | : | | | | :
QY 61 SVTATNLSYNVIOP-GOSTTFGFNGSYSGTNTAPTLTCTAS 101
   | : | | : | | : | : | | | | :
Db 311 QVL-----IARPNGNNGWGVTIQHNGNWTPVTCTAN 344

RESULT 11
US-10-092-880-4
Sequence 4, Application US/10092880
Patent No. US20020164354A1
GENERAL INFORMATION:
APPLICANT: Barenkamp, Stephen J.
TITLE OF INVENTION: HIGH MOLECULAR WEIGHT SURFACE PROTEINS OF NON-TYPEABLE
FILE REFERENCE:
CURRENT APPLICATION NUMBER: US/10/092,880
CURRENT FILING DATE: 2002-03-08
PRIOR APPLICATION NUMBER: 09/155,614
PRIOR FILING DATE: 1998-09-30
PRIOR APPLICATION NUMBER: 08/617,697
PRIOR FILING DATE: 1996-04-01
PRIOR APPLICATION NUMBER: PCT/US97/04707
PRIOR FILING DATE: 1997-04-01
NUMBER OF SEQ ID NOS: 11
SOFTWARE: PatentIn ver. 2.1
SEQ ID NO 4
LENGTH: 1477
TYPE: PR

```



```
; ORGANISM: P. Yoeli
; US-09-823-936-16

Query Match      14.0% Score 76: DB 10: Length 480:
Best Local Similarity 30.0% Pred No. 2.1:
Matches 36; Conservative 5; Mismatches 43; Indels 36; Gaps 6:

Oy 3 GCRATYVYVNSMGSGFTA-----TVTVTNTGSRATSGMTVAWSFG----GNQTVTNW 51
   || || | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 361 GCCATAAGGAGCGGATATAATTTAATAGTACGAATTGATCAGAGGAACCGGATCTT-- 418

Oy 52 NTALTQSGASVTATNLNYSNNVIOQGQSTTFGFGNSYS-----GTNTAPTLT-----CTAS 101
   || | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 419 ATAATCGGAGCTCT-----TTCGAGAGCTTCTTCTGCTTTGGTTTGGACCTCT 465

RESULT 14
US-10-052-586-581
; Sequence 581, Application US/10052586
; Patent No. US20020127584A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Chen, Jian
; APPLICANT: Desnoyers, Luc
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Pan, James
; APPLICANT: Smith, Victoria
; APPLICANT: Watanabe, Colin K.
; APPLICANT: Wood, William I.
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3430R1C1
; CURRENT APPLICATION NUMBER: US/10/052,586
; CURRENT FILING DATE: 2002-01-15
; PRIOR APPLICATION NUMBER: 60/059263
; PRIOR FILING DATE: 1997-09-18
; PRIOR APPLICATION NUMBER: 60/059266
; PRIOR FILING DATE: 1997-09-18
; PRIOR APPLICATION NUMBER: 60/062250
; PRIOR FILING DATE: 1997-10-17
; PRIOR APPLICATION NUMBER: 60/063120
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: 60/063121
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: 60/063486
; PRIOR FILING DATE: 1997-10-21
; PRIOR APPLICATION NUMBER: 60/063540
; PRIOR FILING DATE: 1997-10-28
; PRIOR APPLICATION NUMBER: 60/063541
; PRIOR FILING DATE: 1997-10-28
; PRIOR APPLICATION NUMBER: 60/063544
; PRIOR FILING DATE: 1997-10-28
; PRIOR APPLICATION NUMBER: 60/063564
; PRIOR FILING DATE: 1997-10-28
; PRIOR APPLICATION NUMBER: 60/063734
; PRIOR FILING DATE: 1997-10-29
; PRIOR APPLICATION NUMBER: 60/063870
; PRIOR FILING DATE: 1997-10-31
; PRIOR APPLICATION NUMBER: 60/064103
; PRIOR FILING DATE: 1997-10-31
; PRIOR APPLICATION NUMBER: 60/065311
; PRIOR FILING DATE: 1997-11-13
; PRIOR APPLICATION NUMBER: 60/066120
; PRIOR FILING DATE: 1997-11-21
; PRIOR APPLICATION NUMBER: 60/066466
; PRIOR FILING DATE: 1997-11-24
; PRIOR APPLICATION NUMBER: 60/066772
; PRIOR FILING DATE: 1997-11-24
; PRIOR APPLICATION NUMBER: 60/069335
; PRIOR FILING DATE: 1997-12-11

; PRIOR APPLICATION NUMBER: 60/069425
; PRIOR FILING DATE: 1997-12-12
; PRIOR APPLICATION NUMBER: 60/069870
; PRIOR FILING DATE: 1997-12-17
; PRIOR APPLICATION NUMBER: 60/068017
; PRIOR FILING DATE: 1997-12-18
; PRIOR APPLICATION NUMBER: 60/077450
; PRIOR FILING DATE: 1998-03-10
; PRIOR APPLICATION NUMBER: 60/077632
; PRIOR FILING DATE: 1998-03-11
; PRIOR APPLICATION NUMBER: 60/077649
; PRIOR FILING DATE: 1998-03-11
; PRIOR APPLICATION NUMBER: 60/078886
; PRIOR FILING DATE: 1998-03-20
; PRIOR APPLICATION NUMBER: 60/078939
; PRIOR FILING DATE: 1998-03-20
; PRIOR APPLICATION NUMBER: 60/079664
; PRIOR FILING DATE: 1998-03-27
; PRIOR APPLICATION NUMBER: 60/079786
; PRIOR FILING DATE: 1998-03-27
; PRIOR APPLICATION NUMBER: 60/080107
; PRIOR FILING DATE: 1998-03-31
; PRIOR APPLICATION NUMBER: 60/080194
; PRIOR FILING DATE: 1998-03-31
; PRIOR APPLICATION NUMBER: 60/080327
; PRIOR FILING DATE: 1998-04-01
; PRIOR APPLICATION NUMBER: 60/080333
; PRIOR FILING DATE: 1998-04-01
; PRIOR APPLICATION NUMBER: 60/081049
; PRIOR FILING DATE: 1998-04-08
; PRIOR APPLICATION NUMBER: 60/081070
; PRIOR FILING DATE: 1998-04-08
; PRIOR APPLICATION NUMBER: 60/081195
; PRIOR FILING DATE: 1998-04-09
; PRIOR APPLICATION NUMBER: 60/081838
; PRIOR FILING DATE: 1998-04-15
; PRIOR APPLICATION NUMBER: 60/082568
; PRIOR FILING DATE: 1998-04-21
; PRIOR APPLICATION NUMBER: 60/082569
; PRIOR FILING DATE: 1998-04-21
; PRIOR APPLICATION NUMBER: 60/082704
; PRIOR FILING DATE: 1998-04-22
; PRIOR APPLICATION NUMBER: 60/082797
; PRIOR FILING DATE: 1998-04-22
; PRIOR APPLICATION NUMBER: 60/083322
; PRIOR FILING DATE: 1998-04-28
; PRIOR APPLICATION NUMBER: 60/083495
; PRIOR FILING DATE: 1998-04-29
; PRIOR APPLICATION NUMBER: 60/083496
; PRIOR FILING DATE: 1998-04-29
; PRIOR APPLICATION NUMBER: 60/083499
; PRIOR FILING DATE: 1998-04-29
; PRIOR APPLICATION NUMBER: 60/083559
; PRIOR FILING DATE: 1998-04-29
; PRIOR APPLICATION NUMBER: 60/084366
; PRIOR FILING DATE: 1998-05-05
; PRIOR APPLICATION NUMBER: 60/084414
; PRIOR FILING DATE: 1998-05-06
; PRIOR APPLICATION NUMBER: 60/084639
; PRIOR FILING DATE: 1998-05-07
; PRIOR APPLICATION NUMBER: 60/084640
; PRIOR FILING DATE: 1998-05-07
; PRIOR APPLICATION NUMBER: 60/084643
; PRIOR FILING DATE: 1998-05-07
; PRIOR APPLICATION NUMBER: 60/085573
; PRIOR FILING DATE: 1998-05-15
; PRIOR APPLICATION NUMBER: 60/085579
; PRIOR FILING DATE: 1998-05-15
; PRIOR APPLICATION NUMBER: 60/085580
; PRIOR FILING DATE: 1998-05-15
; PRIOR APPLICATION NUMBER: 60/085582
; PRIOR FILING DATE: 1998-05-15
; PRIOR APPLICATION NUMBER: 60/085700
```

PROBES USEFUL. FO

PRIOR FILING DATE: 1998-05-15	
PRIOR APPLICATION NUMBER: 60/086023	
PRIOR FILING DATE: 1998-05-18	
PRIOR APPLICATION NUMBER: 60/086392	
PRIOR FILING DATE: 1998-05-22	
PRIOR APPLICATION NUMBER: 60/086486	
PRIOR FILING DATE: 1998-05-22	
PRIOR APPLICATION NUMBER: 60/087098	
PRIOR FILING DATE: 1998-05-28	
PRIOR APPLICATION NUMBER: 60/087208	
PRIOR FILING DATE: 1998-05-28	
PRIOR APPLICATION NUMBER: 60/087609	
PRIOR FILING DATE: 1998-06-02	
PRIOR APPLICATION NUMBER: 60/087759	
PRIOR FILING DATE: 1998-06-02	
PRIOR APPLICATION NUMBER: 60/087827	
PRIOR FILING DATE: 1998-06-03	
PRIOR APPLICATION NUMBER: 60/088025	
PRIOR FILING DATE: 1998-06-04	
PRIOR APPLICATION NUMBER: 60/088028	
PRIOR FILING DATE: 1998-06-04	
PRIOR APPLICATION NUMBER: 60/088029	
PRIOR FILING DATE: 1998-06-04	
PRIOR APPLICATION NUMBER: 60/088033	
PRIOR FILING DATE: 1998-06-04	
PRIOR APPLICATION NUMBER: 60/088167	
PRIOR FILING DATE: 1998-06-05	
PRIOR APPLICATION NUMBER: 60/088202	
PRIOR FILING DATE: 1998-06-05	
PRIOR APPLICATION NUMBER: 60/088212	
PRIOR FILING DATE: 1998-06-05	
PRIOR APPLICATION NUMBER: 60/088217	
PRIOR FILING DATE: 1998-06-05	
PRIOR APPLICATION NUMBER: 60/088326	
PRIOR FILING DATE: 1998-06-04	
PRIOR APPLICATION NUMBER: 60/088655	
PRIOR FILING DATE: 1998-06-09	
PRIOR APPLICATION NUMBER: 60/088722	
PRIOR FILING DATE: 1998-06-10	
PRIOR APPLICATION NUMBER: 60/088738	
PRIOR FILING DATE: 1998-06-10	
PRIOR APPLICATION NUMBER: 60/088740	
PRIOR FILING DATE: 1998-06-10	
PRIOR APPLICATION NUMBER: 60/088811	
PRIOR FILING DATE: 1998-06-10	
PRIOR APPLICATION NUMBER: 60/088824	
PRIOR FILING DATE: 1998-06-10	
PRIOR APPLICATION NUMBER: 60/088825	
PRIOR FILING DATE: 1998-08-10	
PRIOR APPLICATION NUMBER: 60/088826	
PRIOR FILING DATE: 1998-06-10	
PRIOR APPLICATION NUMBER: 60/088861	
PRIOR FILING DATE: 1998-06-11	
PRIOR APPLICATION NUMBER: 60/088863	
PRIOR FILING DATE: 1998-06-11	
PRIOR APPLICATION NUMBER: 60/088876	
PRIOR FILING DATE: 1998-06-11	
PRIOR APPLICATION NUMBER: 60/089090	
PRIOR FILING DATE: 1998-06-12	
PRIOR APPLICATION NUMBER: 60/089105	
PRIOR FILING DATE: 1998-06-12	
PRIOR APPLICATION NUMBER: 60/089512	
PRIOR FILING DATE: 1998-06-16	
PRIOR APPLICATION NUMBER: 60/089514	
PRIOR FILING DATE: 1998-06-16	
PRIOR APPLICATION NUMBER: 60/089538	
PRIOR FILING DATE: 1998-06-17	
PRIOR APPLICATION NUMBER: 60/089598	
PRIOR FILING DATE: 1998-06-17	
PRIOR APPLICATION NUMBER: 60/089653	
PRIOR FILING DATE: 1998-06-17	
PRIOR APPLICATION NUMBER: 60/089908	

[illegible]

US-09-864-761-36047

Patent NO. US20020100000

APPLICANT: Rank, David R.

APPLICANT: Chen, Wensheng

TITLE OF INVENTION: GENE

CURRENT APPLICATION NUMBER

: PRIOR APPLICATION NUMBER: 2000 02

2000-05

;
; PRIOR APPLICATION NUMBER: 2000-09

; PRIOR APPLICATION NUMBER: 2000-19
 ; FILING DATE: 2000-19

PRIOR APPLICATION NUMBER: 2000-09

PRIOR FILING DATE: 2001-0

PRIOR FILING DATE: 2001-0

PRIORITY FILING DATE: 2001-0

PRIOR FILING DATE: 2001-0

PRIOR FILING DATE: 2001-0

PRIOR FILING DATE: 2001-09-11

PRIOR FILING DATE: 2001-
PRIORITY NUMBER

; PRIOR FILING DATE: 2001
; REGISTRATION NUMBER:

PRIOR FILING DATE: 2001-01-01
PUBLICATION NUMBER: 2002-01-01

;; PRIOR FILING DATE: 1998-01-01
;; PRIOR APPLICATION NUMBER: 08/100,000

;; PRIOR FILING DATE
;; PRIOR APPLICATION NUMBER

;	PRIOR FILING DATE	PRIOR APPLICATION NUMBER
;		
;		

NUMBER OF SEQ ID NOS: 49

SEO ID NO 36047

TYPE: PRT

FEATURE:

; OTHER INFORMATION: EXPD
; INFORMATION: EXPD

OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 1.1
OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.5
OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.5
OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 1.1
OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 2
OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 4.7
OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 2.1
OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 1.4
US-09-864-761-36047

Query Match 13.5% Score 73.5; DB 10; Length 688;
Best Local Similarity 32.0%; Pred. No. 5.5;
Matches 32; Conservative 12; Mismatches 43; Indels 13; Gaps 4;
QY 3 GCRATYVNVSDMGSGFTATVTVTGSRATSGWTVAWSEFGNGQVTNYYNTALTQSGASV 62
Db 460 GSEATTVSTTS-----SETTTASTTGSEMT---TVFTTVSETTVSTIGSEATTSSAAGS 511
QY 63 TATNLSYNNVLOPGOSTFGNGSYSGT-NTAPLTCTAS 101
Db 512 EATTTS-----TEGSETTTASTAGSETTTASTAGSETTTAS 547

Search completed: November 13, 2002, 12:08:37
Job time : 2.95905 secs

